Department of Physics and Astronomy, University of Canterbury,
Private Bag 4800, Christchurch, New Zealand

Validation of Deformable Image Registration for Head & Neck Cancer Adaptive Radiotherapy

Ihab Ramadaan, BTech (Hons)

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in the University of Canterbury

Supervisor: Dr. Jürgen Meyer

Co-Supervisor: Mr. Karsten Peick

April, 2013

Abstract

Anatomical changes can have significant clinical impact during head and neck radiotherapy. Adaptive radiotherapy (ART) may be applied to account for such changes. Implementation of ART to alter dose delivery requires deformable image registration (DIR) to assess 3D deformations. This study evaluates the performance and accuracy of a commercial DIR system for clinical applications.

The investigations in this project were carried out using images of induced changes in two standard radiotherapy phantoms (RANDO® and CIRS®) and one in-house built phantom. CT image data before and after deformation of the phantoms were processed using Eclipse / SmartAdapt® v.10 system employing a Demons-based algorithm. A DIR protocol was designed, and algorithm performance was assessed quantitatively, using volume analysis and the Dice Similarity Index (DSI), and also evaluated qualitatively. In addition, algorithm performance was assessed for 5 head and neck cancer patients using clinical CT images. Each original planning CT image containing contours of 10 volumes of interest including treatment target volumes and organs at risk was deformed to match a second CT image acquired during the course of the treatment. The original structures were deformed, copied onto the target image and compared to reference contours drawn by 3 radiation oncologists.

Phantom investigations gave varied results with average DSI scores ranging from 0.69 to 0.93, with an overall average of 0.86 ± 0.08. These quantitative results were reflected qualitatively, with generally accurate matching between reference and DIR-generated structures. Although air gaps in the phantoms compromised algorithm performance and gave rise to physically aberrant results. Clinical results were generally better with a DSI range of 0.75-0.99 and an overall average of 0.89 ± 0.05, suggesting high DIR accuracy. Qualitatively, some minor contour deformations were noted, as well as artefacts in the axial direction that were due to the CT slice resolution (3 mm) that was used to scan the patients. In addition, contour propagation between images using DIR reduced the time required by physicians to contour the images of head and neck cancer patients by ~47%.

This study demonstrated that deformable image registration using a Modified Demons algorithm yields clinically acceptable results and time-saving benefits in contouring that improve clinical workflow. The study also showed that it is feasible to incorporate deformable image registration as part of an adaptive radiotherapy strategy for head and neck cancer, provided further studies are designed to carry out accurate and verifiable dose deformation.

Acknowledgments

Firstly I would like to thank my Clinical Supervisor Karsten Peick for his encouragement and enthusiasm in helping me complete this project and for providing his clinical insight throughout the course of the study – thanks KP. Tremendous thanks also go to my Academic Supervisor Dr. Juergen Meyer for his wholehearted and involved support during the first year of the project, and his continued advice after leaving the university.

I would like to acknowledge the organisational support from Lynne Greig, Chief Physicist at Wellington hospital's Blood and Cancer Centre, and her efforts in providing me with the time and resources needed to complete this project. A big thank you also goes to Wellington's Radiation Oncologists Drs. David Hamilton, Doug Iupati, and James Evans for taking the time to check and modify the DIR-generated contours in the clinical studies.

My thanks and appreciation to my parents and brothers for their faith and inspiration throughout cannot be understated. Their confidence and encouragement were a boon when things were at their most stressful.

Finally and chiefly, I extend a special acknowledgment and thanks to Dr. Rob Louwe, surrogate supervisor and Medical Physicist extraordinaire, for the expert knowledge and advice he contributed in helping me complete this thesis. Thank you Rob, I am very grateful for the assistance and copious amounts of your own time that you provided for this project.

Table of Contents

Glossary

ART	*Adaptive Radiotherapy*
C3	*Third cervical vertebra*
CT	*Computed x-ray tomography*
CTV	*Clinical target volume*
D+ER	*Fast Demons algorithm with elastic regularisation*
DIR	*Deformable image registration*
DSI	*Dice Similarity Index*
DVF	*Deformation vector field*
FF	*Free form algorithm with calculus of variance*
GTV	*Gross tumour volume*
HS	*Horn-Schunck optical flow algorithm*
HS+LK	*Combined Lucas-Kanad and Horn-Schunck optical flow algoirhtm*
OAR	*Organ at risk*
PTV	*Planning target volume*
RO	*Radiation oncologist*
ROI	*Region of interest*
SA	*SmartAdapt deformable image registration programme*
SM	*Similarity measure*
TPS	*Treatment planning system*
TV	*Target volume*
VOI	*Volume of interest*

1. Introduction

1.1. Background

Head and neck cancer categorises a group of primary tumours and their associated metastases (secondary breakaway tumours) that afflict the throat, nose, mouth, salivary glands and cervical (neck) lymph nodes, but excludes malignancies of the brain [1]. Different treatment strategies are employed clinically, including surgery, chemotherapy and radiotherapy. The work in this project is tied to the imaging aspect of radiotherapy, although the clinical cases studied here are of patients who had undergone surgery and/or chemotherapy in addition to radiotherapy as part of the overall treatment strategy.

In radiation treatment of head and neck cancer, medical images in a variety of modalities are used to design a treatment plan for the delivery of the radiation dose prescribed by the Radiation Oncologist (RO), a physician specialised in the treatment of cancer patients with radiation. Medical images are used to delineate patient anatomy and identify the extent and location of the malignancies [2]. The anatomy of the patient is contoured using a computerised treatment planning system (TPS) into several volumes of interest (VOIs) that delineate either target volumes (TVs), which are the objects of the treatment, or organs at risk (OARs) that are structures to be spared from irradiation [3].

The principal imaging modality in radiotherapy planning is x-ray computed tomography (CT) because it serves the double role of providing good anatomical image resolution and contrast, as well as the tissue density information required for determining the radiation dose to be delivered to the patient [2, 4]. However, the CT images used for treatment planning constitute only a snapshot in time of the patient's anatomy at the beginning of radiotherapy, which

commonly involves 25-35 fractions delivered in a time span of several weeks. During this period, the patient may undergo significant anatomical changes.

Throughout the course of radiotherapy, patients often exhibit marked changes in the anatomical position, shape and volume of TVs and OARs. These anatomical changes may originate from various sources, such as day-to-day variation in positioning of the patient on the treatment couch, weight change, inflammation, or tumour shrinkage. These changes can have a significant impact on the clinical outcome of head and neck treatments due to the proximity and often overlapping organisation of TVs and OARs [5]. Consequently, such changes can result in a reduced dose actually delivered to the TV than prescribed, or a higher dose to OARs than accepted during the treatment preparation phase. The challenge for treatment optimisation presented by these facts led to the promulgation of adaptive radiotherapy (ART) in 1997 by Di Yan *et al.* in their seminal paper 'Adaptive Radiation Therapy' [6]. Therein, ART is defined as a radiation treatment process where the treatment plan can be modified in the course of treatment using a systematic feedback of measurements. To facilitate this process, regular imaging is needed for up-to-date information on the condition of the patient's anatomy as the treatment progresses. These additional images are registered (superimposed/matched) to the original planning CT scan (pCT) used for treatment planning, in order to determine the extent of any anatomical changes. Careful analysis of the registered images may then indicate a modification of the treatment plan.

The registration of medical images can be a complex process, especially when deformable image registration (DIR) is involved. Currently, rigid image registration (RIR) is ubiquitously used in radiotherapy departments to match clinical images acquired at different times, e.g. before and during treatment. This type of registration treats the images as rigid bodies so matching is limited to the translation and/or rotation of one image to correspond to the other.

RIR usually uses bony anatomy or fiducial markers to register the images because they define landmarks or frames of reference for translation or rotation. However, the various TVs and OARs can differentially move or change in shape and size between images so that translation and/or rotation of the whole image is insufficient to account for all such changes. DIR on the other hand uses a computer algorithm to relate the points in one image to corresponding points in the other by warping the first image to match the second [7]. Essentially, this method treats the images as non-rigid bodies that are capable of heterogeneous deformations in addition to translation and rotation, and consequently every pixel/voxel of an image can be manipulated differentially to account for all anatomical variations. However, there are limitations and conditions associated with this concept and it is the primary topic of this project that will be explored in detail in Chapter 2.

The next logical step in ART is to determine the impact that anatomical changes during treatment can have on the cumulative dose delivered to the patient, which is necessary to avoid under-dosing TVs and/or over-dosing OARs [8]. To do this, the results of DIR can be used to deform the dose distribution as calculated by the TPS. Alternatively, daily imaging can be used to accumulate the dose from each treatment fraction to determine the dose received by a given VOI [9-11]. However, before clinical application of this type of ART, the accuracy of DIR results needs to be established. Therefore, objective methods of evaluating DIR accuracy are required [12, 13]. Until recently, it was difficult to validate the accuracy of deformed dose distributions that are based on image manipulation without direct measurement; specifically, accounting for the dose to missing tissues (e.g. tumour shrinkage) that take the dose with them. Therefore, if the deformation vector field obtained from image deformation is to be used to warp dose, then specific tools and methods are required to verify the deformed dose; for example direct dose measurements in deformable gel phantoms [14]. However, the subject of dose deformation/accumulation is a substantial area of research and

is beyond scope of this study, which is concerned with the investigation of DIR accuracy from the perspective of imaging only and does not extend to dose deformation.

Potential applications of DIR are not limited to ART. DIR can also be used in computerised contouring and auto-segmentation of physical volumes in medical images where contour volume propagation across multiple images is applied using anatomical ‘atlases’ as a starting point [15, 16]. This application of DIR can have substantial benefits on time-saving and easing of clinical workflow [17].

1.2. Purpose

The purpose of this project is to validate a commercial deformable image registration programme for clinical use in head and neck cancer radiotherapy. The aim is to evaluate the accuracy and performance of DIR by quantitative and qualitative means using both clinical images, and images of simulated changes in phantoms. The experimental part of this project will be divided in two parts: (I) investigations using phantoms in which simple and known changes are induced to assess the limitations and boundary conditions of the DIR algorithm(s). (II) Clinical case studies using patient images in treatments where multiple CT scans were acquired. Quantitative assessment of DIR performance will be carried out using parameters that measure correlation of images and VOIs called similarity measures (SMs), while qualitative assessment of DIR performance will be made by examining CT images and the change in contours associated with several VOIs. Based on these results, the feasibility of clinical application of the DIR programme will be discussed, as well as the potential applications of DIR to improve clinical workflow. Furthermore, the future steps needed before implementation of DIR as part of an ART strategy for head and neck cancer will be reviewed.

2. Deformable Image Registration

Safe and effective patient treatment in radiotherapy depends on accurate delivery that requires detailed information about patient anatomy. Medical images delineate patient anatomy when the different organs and structures are at a given position, size and shape within the patient. In radiotherapy planning, the clinician would delineate different regions of these images into target volumes (TVs) and organs at risk (OARs). The images can be matched to each other in what is known as image registration to track any changes in these structures that have occurred in the intervening period between when each image was taken.

Image registration calculates a transformation function that correlates the points in one image of an object to their counterparts in the other image of said object [18]. Currently, rigid image registration is the *modus operandi* of image registration software in clinical departments [19]. However, rigid image registration only allows the manipulation of images to 6 degrees of freedom. That is, the images can only be translated along the 3 cardinal axes – sagittal (x), coronal (y) and axial (z) – and rotated about these axes: pitch, yaw and roll respectively. Thus, when structures change independently to each other (e.g. tumour shrinkage), rigid registration cannot account for all anatomical changes in an image simultaneously, and so extra margins are added to the TVs and OARs to ensure coverage and sparing, respectively. To precisely evaluate all anatomical changes at the same time, it is possible to use deformable image registration (DIR), provided it is preceded by a rigid registration to ensure minimal calculation time and accurate registration results.

It is important to point out that the reason that DIR has not been widely performed clinically is due to the difficulty in its implementation and validation [20]. In general image registration programmes (rigid or deformable) allow for manual corrections of the results, because the human brain processes gray scale images so much better than a computer [12, 21]. However,

this is not viable for DIR because the optimal registration currently cannot be objectively defined as there are no objective metrics or gold standard with which to validate the accuracy of a deformed image. Nonetheless, continuing research into algorithm design and optimisation seeks to characterise the suitability of different methods of DIR using more sophisticated phantoms and artificially warped clinical images [22-24].

2.1. Mechanism of DIR

Firstly, it is important to clarify some of the terminology associated with DIR. In this context, each image of the image pair undergoing a registration is identified by the monikers of 'source' and 'target'. It may help to visualise that the source image is the origin of the deformation data and the target image is its destination. Another common convention used in DIR research is to name the source and target image as the 'moving' and 'fixed' image, respectively. In this study, the source/target naming convention will be used.

DIR employs algorithms that allow the alignment of non-linearly or non-uniformly mismatched datasets. This is normally achieved by creating a deformation vector field (DVF) that defines a voxel-to-voxel mapping between the source and target images [7, 18]. This mapping is typically formulated as an optimisation problem where the voxel mapping (solution) is found by maximising the similarity (i.e. mutual information) between the 2 images [25]. There are two main categories of DIR methods: point-based methods and voxel intensity-based methods [26, 27].

2.1.1. Point-based DIR Methods

Point-based techniques first minimise the distance between features or landmarks (e.g. points or surfaces) of corresponding anatomical structures and match the rest of the data using an interpolation technique such as a radial basis function and thin-plate splines [28]. These landmarks need to be identified and matched on both source and target images, which require a certain amount of human interactions. Furthermore, these methods are based on physical models to follow anatomical changes that predicate upon possessing prior knowledge of the material properties and heavy computation demands [18, 26]. Figure 2.1 gives a schematic of point-based DIR, where user-specified landmarks are placed on easily identifiable anatomy, that can be clearly seen on both images. The objective of the DIR algorithm would be to match all these points as closely as possible as a first priority, which is then followed by the deformation of the rest of the source image according to the physical model input into the algorithm.

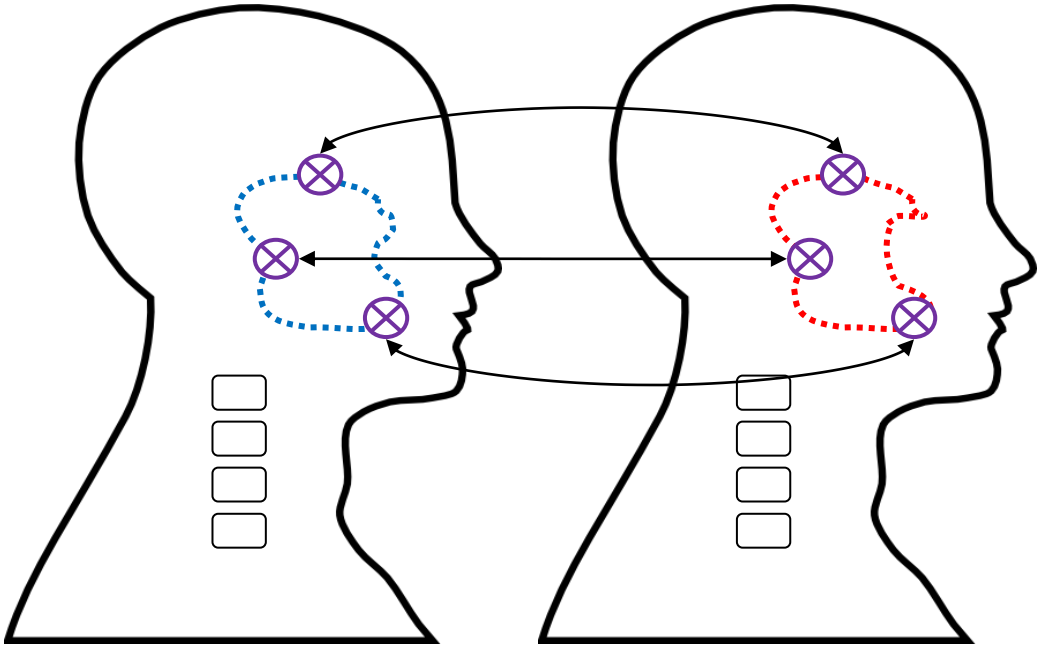

Figure 2.1: Point-based deformable registration techniques. The original source image (left) and the target image (right) are marked by user-defined reference points (purple crossed circle) that are matched first. Blue and red dashed contours represent a volume of interest undergoing a shape and volume change.

A desirable DIR strategy should be fast, accurate, fully automatic and able to handle large deformation volumes. Point-based techniques are not automatic, some are time-consuming because they require manually selecting a large number of landmarks, which increases the likelihood of the user making an error as they must accurately label both images with landmarks [29]. However, some studies have introduced ways of overcoming these limitations, such as analysis of variance based validation and landmark weighting [26, 30], which are beyond the scope of this study.

2.1.2. Voxel Intensity-based Methods

In comparison, voxel intensity-based methods use similarity measures between the images such as the root mean square error or mutual information (Section 2.3), and attempt to minimise the former parameter and maximise the latter parameter when deforming the source image to match the target image as closely as possible. These techniques use all the voxels in the image to deform it to satisfy a global objective/cost function, i.e. the images being registered are matched voxel-by-voxel. Furthermore, algorithms based on this method also have further local objectives built-in to enhance performance, such as smoothness constraints [29]. For example, a local objective may be added to an algorithm that provides relatively smooth or continuous deformations in regions of the image that have predictable deformations based on *a-priori* knowledge of the anatomic region displayed by the image [31]. Figure 2.2 illustrates a schematic of intensity-based DIR techniques, where groups of voxels in the source image are matched to corresponding groups of voxels in the target image that possess identical or similar intensities or 'grey values', which in turn is based on several aspects of image quality like resolving power and image reproducibility for a given patient. In this illustrated example, the grey-highlighted voxels possess intensities based on the anatomy they contain. So when the volume of interest (VOI) deforms from its original shape (blue) to its final shape (red) the voxel intensities would appear to have 'moved'. It is important to conceptually appreciate that the actual voxels, being virtual representations of

elements of the CT detector, do not actually move but that the grey value of interest along with the anatomy it represents moves from one voxel to another.

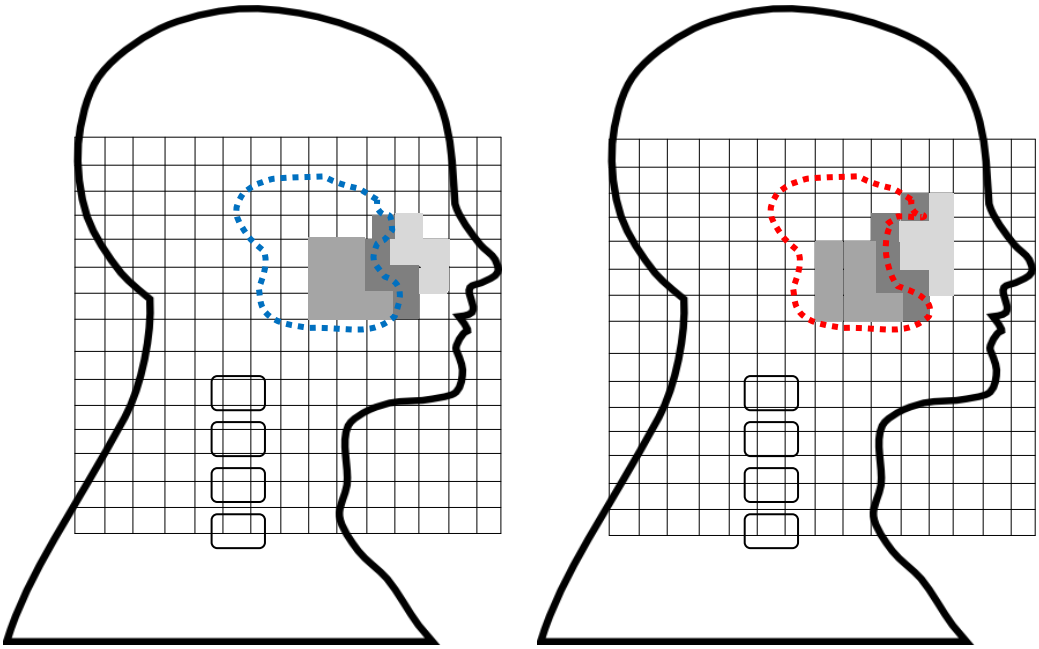

Figure 2.2: Intensity-based deformable registration techniques. The transparent grid represents all the voxels in the DIR region of interest. The gray voxel group has a particular intensity distribution representative of the anatomy it contains.

Intensity-based methods have the advantage over point-based techniques in that they do not require any feature extraction process, and therefore are often faster and fully automatic [18]. However, the disadvantages of such methods is that they only seek to match voxel intensities without the guidance of a physical model or constraint of the contour, i.e. if neighbouring voxels have similar intensities to those included in a contour, they may be wrongly incorporated in the contour. In addition, matching contrast-enhanced CT images with non-enhanced CTs can also affect algorithm performance, since the same tissues can have different voxel intensities.

2.2. DIR Algorithms

Due to the advantages mentioned in Section 2.1 and despite the disadvantages, all the algorithms used in this project and discussed hereafter are voxel intensity-based techniques and no point-based algorithms will be discussed. Specifically, the main advantage of intensity-based methods, automation, precludes the requirement of point-based method for expert knowledge of anatomy for the purpose of feature/landmark selection. There are many intensity-based DIR algorithms currently used in medical imaging that have their origins in thermodynamic theory and computer science [32, 33]. Most algorithms operate by solving partial differential equations governing the direction of registration, and smoothness constraints [18]. The Optical Flow Method and the Demons method are the 2 principal algorithms used in this study. They share a common approach to calculating the DVF and optimising DIR mapping.

This DIR process has been described in detail by Yang & El Naqa [34]. A general explanation of the common approach of the two methods is given here. Consider the registration of a source and target image with intensity distributions *I* and *J*, respectively, each defined in its own domain. The DIR process attempts to compute the DVF, *V*, in order to optimise the system energy (*E*) equation by maximising the similarity function (*S*) while maintaining the smoothness constraint (*R*) boundaries:

$$E = \int_{\Omega} S(V(I), J)\, d\Omega + \alpha^2 \int_{\Omega} R(V)\ d\Omega \qquad (2.1)$$

where *V(I)* is the deformed source image, Ω is the image domain, and α is a constant used for weighting *R* [34]. *E* corresponds to the voxel intensity distribution of the registration, and *S* describes how *V* relates the voxels in *I* to the voxels in *J*. *R* is the function that controls the extent to which the intensity can change.

The DVF is defined on the co-ordinates of *J*, i.e. *V* is the transformation vector field yielding the same array dimension as *J*. Each element of *V* is a 3D vector, associated with a voxel in *J*, and defines the geometric mapping of *I* to *J*.

2.2.1. Algorithm Tested in Initial Investigations

The fast free-form (FF) DIR registration algorithm is a variation of the same concept presented in the previous section. It was only used in the Initial Investigations of this study (Section 3.3.1) and for this reason will only be introduced briefly as a detailed explanation of its mechanism is beyond the scope of this project.

Fast Free-Form Deformable Registration via Calculus of Variations

The FF algorithm works to minimise an energy function in combination with smoothness and similarity measures in similar fashion to that described in Equation (2.1) [35]. The minimisation problem is represented as a set of partial differential equations that are solved iteratively. This solution can be considered as finding a compromise between the similarity measure *S* and the smoothness constraint *R*.

2.2.2. Optical Flow – Horn-Schunck Method

The Horn-Schunck Optical Flow (HS) algorithm was used in part of the Phantom Studies component of this project (Section 3.3.2). The concept of 'optical flow' was originally introduced to image registration from research into the application of motion capture cameras for artificial intelligence studies [33]. Optical flow is the distribution of apparent velocities of movement of brightness patterns in an image. It can arise from the motion of objects between image frames, and gives information about the spatial arrangement of the objects under study and the rate of change of their arrangement, while discontinuities in the optical flow can be used for image segmentation [33].

The optical flow method of image matching is designed to find small deformations in temporal sequences of images. In its original form, optical flow tracks the same object in 2 successive frames taken in 2 successive instances in time, and so the difference in how the object appears in each frame is seen as a movement over time, i.e. a velocity. But, for medical images that are not necessarily successive, it is more general to consider the velocity is just a displacement [32]. Nonetheless, the basic requirement for small changes still holds, so essentially, it assumes that changes occur gradually and in a continuous manner so that the apparent displacement of a grey value from one image to the next is small [32, 33].

For the purposes of DIR, optical flow can be calculated by computing the instantaneous displacement for every point (voxel) in the target image from the position of its counterpart in the source image [32, 33]. Obtaining such a displacement vector at every voxel of the target image will give a field of displacement vectors, otherwise known as a deformation vector field, DVF.

In the implementation of this algorithm into an automatic process, Yang & El Naqa [34] use a 'multiple pass' approach developed by Iu & Lin that builds on the HS algorithm [36]. This approach entails performing the registration multiple times whereby successive iterations are computed based on the result of the previous pass.

2.2.3. Modified Demons DIR Algorithm

The principal algorithm used in this project is a variant by Wang *et al.* [37] of the Demons algorithm originally put forth by Thirion [32]. The original Demons algorithm used the concept of diffusing models to perform image-to-image matching. It considers the object boundaries in the target image as semi-permeable membranes and that the source image is a deformable grid model that diffuses through these interfaces, by action of effectors or

'demons' situated within the membranes (Figure 2.3) – an analogy of Maxwell's demons in thermodynamic theory [32].

More specifically, the algorithm uses intensity gradient information from a target image to determine the 'demons' force required to deform the source image. That is, the diffusing model assumes that the local demons at every voxel location are applying invisible 'forces' that push the voxels of the source image into matching up with the target image, which may not be efficient if the gradient on the target image is low [37].

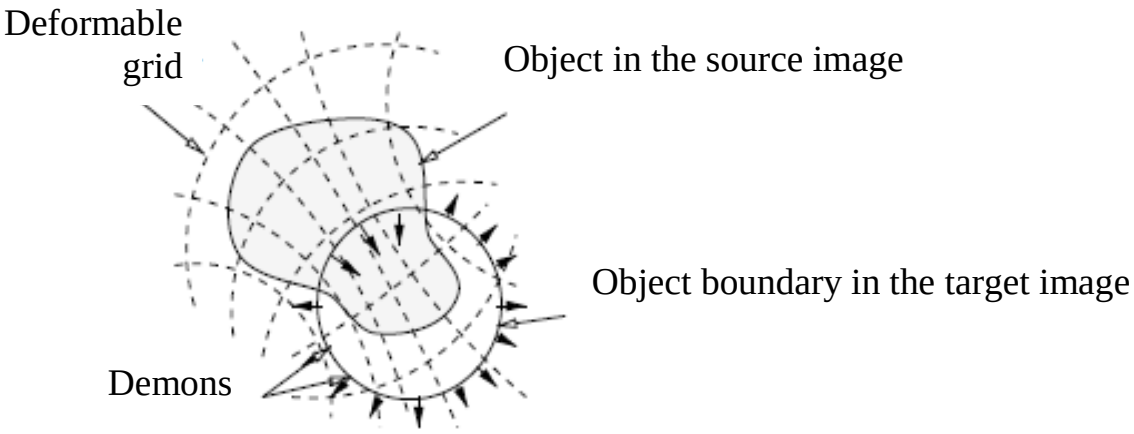

Figure 2.3: Diagram from Thirion illustrating the diffusing models where a deformed image, treated as a deformable grid, is diffusing through the contours of the objects in the static image, by the action of effectors, called demons, situated in these surfaces [32].

This diffusion model is related to the concept of 'optical flow' (Section 2.2.2), where the intensity gradient of the target image, represents the relationship between neighbouring points in the target image, so it is an 'internal' force originating from the target image. On the other hand, the displacement vector of a given point or voxel intensity represents a differential force arising from the interaction between the target and the source image. Therefore, it is external to the target image as it also uses information from the source image [32, 37]. This distinction becomes important when Wang *et al.* accelerate the original Demons algorithm by using both internal and external forces rather than only the internal force which does not make use of the information contained in the source image [37].

In the original Demons algorithm the displacement or DVF is calculated iteratively and each repetition is followed by regularisation of the deformation field using a Gaussian filter. The Gaussian filter is applied locally for the computation of each displacement vector, i.e. the filter will use information from a number of neighbouring voxels specified by the size set for the filter. This means the deformation is calculated from local image only so it is essentially a smoothing operation to suppress noise and preserve the geometric continuity of the deformed image [37]. The iterative calculation becomes more computationally complex as the size of the images becomes bigger and as the intensity difference between them is greater, because there is more information to process.

Wang *et al.* modified this algorithm by adding an 'active force' component to the diffusion process. They pointed out that the gradient information that drives the deformation is inefficiently taken from the target image only, and that it can be improved by assuming that diffusion is bi-direction. That is, the demons at any point in image space will not only produce a force that allows a deformable source image to diffuse into a corresponding target image but also produce a force that allows the target image to diffuse into the corresponding source image [37]. The addition of this 'active force' results in the Demons algorithm to converge more quickly and require fewer iterations [37].

Multi-resolution Approach

The main assumption of optical flow and the demons algorithms that the deformations in the registration are small (displacement of a few voxels) is not always true in clinical settings [5, 18, 37]. If the magnitude of deformations is large, many iterations are required to recover the deformation, which leads to a high computational demand. Worse, the optimisation process can get stuck in a locally optimal solution that prevents algorithm convergence and results in an unrealistic deformation [18]. One method to minimize the effects of large deformations is to use a coarse-to-fine multi-resolution approach [37]. This technique not only alleviates the

limitation to small deformations of the single-resolution model but also its computational complexity, thus improving both computation time and algorithm convergence speed [18]. The way it does that is by down-sizing each image, i.e. reducing its resolution, which allows the algorithm to quickly obtain a rough approximation of the large displacements [18, 38].

The multi-resolution algorithm builds an image pyramid with the image in its original resolution as its bottom. Each level of the pyramid consists of the down-sampled version of the image lying at the lower level with a down-sampling factor of 2 [18]. For example, if the original image is 256x256x256 voxels and the pyramid has 3 levels, each level has dimension from bottom to top, 256x256x256, 128x128x128 and 64x64x64, respectively. The registration algorithm is initialised on the coarsest resolution image. After a few iterations, the deformation map is up-sampled to the next finer resolution in order to get the same resolution as the image resolution at this level of the pyramid. Subsequently, the deformation map is applied to the source image and the process is repeated until the finest level is reached, i.e. the resolution of the original image [18]. The regularisation component (the Gaussian filter) is needed to obtain a smooth deformation field, i.e. where neighbouring voxels have similar deformation vectors. This regularisation is performed in all levels of the pyramid.

Lastly, Wang *et al.* added another normalisation factor to their modification of Thirion's original Demons algorithm that allows force strength to be adjusted adaptively in each iteration. A smaller value of this factor is initially used for relatively large deformations and the step size then reduced when the algorithm approaches convergence [37].

In the end, there are 4 adjustable parameters to control the Demons registration process: 1) the number of levels in the pyramid, 2) the number of iterations at each level, 3) the variance of the Gaussian regularisation filter at each level, and 4) the force strength normalisation factor.

2.3. Similarity Measures

Quantitative validation of a DIR algorithm for clinical images tends to be difficult because: (I) there is no gold standard with which to compare the deformed image. (II) Grey scale CT images have discrete grey values so in principle it is not possible to distinguish two voxels that have the same intensity. Therefore, one can only assert that a given voxel on the target image represents the same tissue as another voxel in the source image by correlation or mutual information of its surrounding voxels. In addition, DIR is inherently degenerative since multiple solutions may exist for given match of image intensity [37]. Furthermore, qualitative evaluation of DIR algorithm performance alone is not sufficient to provide comprehensive information on how anatomical changes will affect treatment outcome, especially if an attempt is made at assessing the dosimetric implications for treatment planning of such changes. Therefore, researchers in DIR have proposed many ways to quantitatively evaluate DIR performance, typically referred to as Similarity Measures (SMs), which cover a broad range of quantitative tools and parameters. SMs either compare the registered images and the delineated structures in these images after DIR to their status before DIR, or assess how closely the deformed source image and its structures match the target image. The most common SMs used in DIR include Mutual Information (MI), Entropy (H), Dice Similarity Index (DSI), Pearson's Product-Moment Correlation Co-efficient (*r*), and Root Mean Square Error (RMSE) [18, 34, 39-41], while some newer studies have used other measures such as analysis-of-variance (ANOVA) [30]. In this project, the DSI and *r* were used explicitly in TV and OAR comparisons, and MI was used implicitly in the registration process.

2.3.1. Dice Similarity Index

The Dice Similarity Index (DSI) is a volume-based SM that quantitatively determines the accuracy of DIR by comparing a DIR-propagated VOI contour with a reference VOI contour

on the same image. In practice, after a source image is deformably registered to a target image, the computed DVF is used to deform all VOI contours on the source image accordingly. The deformed VOIs are then copied from the deformed source image unto the target image while the images are still superimposed. The target image would have user-drawn contours of the same VOIs (e.g. GTV, parotid glands) that act as reference VOIs. Ideally, the DIR-propagated contours should accurately and precisely match the VOIs. However, in practice there will be discrepancies between the DIR-generated contours and the user-drawn contours, which can be quantified by the DSI. The DSI indicates the overlapping ratio between the two VOIs being compared [18, 42], it is defined as:

$$DSI = \frac{2 \cdot |A \cap B|}{|A| + |B|} \qquad (2.2)$$

where A and B are the two groups of voxels contained within the VOIs being compared (Figure 2.4). If the two VOIs are perfectly matched, i.e., $2 \cdot |A \cap B| = |A| + |B|$, the DSI will equal 1. Whereas, two disjoint volumes ($|A \cap B| = 0$) lead to a DSI equal to 0.

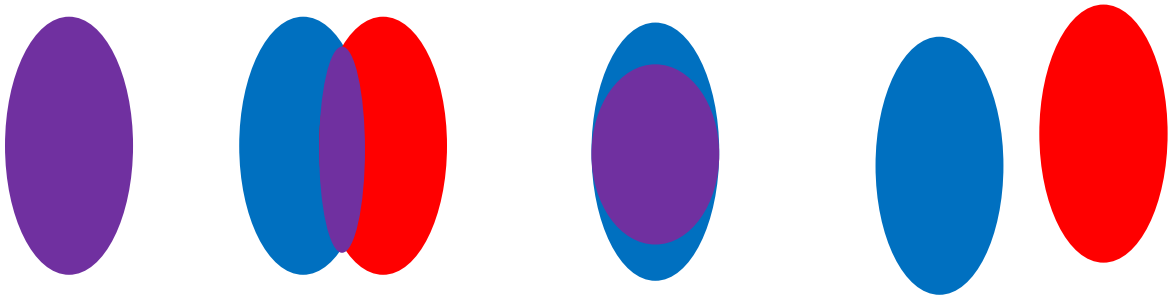

Figure 2.4: Overlapping volume scenarios for DSI calculations. (I) Perfect match between the VOIs, DSI = 1. (II) Partial match, where both VOIs have an equal volume but are relatively displaced, 0 < DSI < 1. (III) Partial match, where both volumes have the same displacement but a mismatch in volume magnitude, 0 < DSI < 1. (IV) Disjoint volumes, no overlap, DSI = 0.

Figure 2.4 demonstrates that an intermediate value for the DSI corresponds to a partial match that may arise from a mismatch in the magnitude of the VOIs, or a mismatch in the displacement of the VOIs, or a combination of both.

2.3.2. Pearson's Correlation Co-efficient

Pearson's product moment correlation co-efficient (*r*) is a voxel intensity-based criterion that was used in the Phantom Studies (Section 3.3) to assess the accuracy of the DIR results. r is defined as:

$$r = \frac{\sum_{i \in A}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i \in A}(x_i - \bar{x})^2}\sqrt{\sum_{i \in A}(y_i - \bar{y})^2}} \qquad (2.3)$$

where x_i is the intensity of the *i*th voxel in the target image and y_i is the intensity of the corresponding voxel in the source image and $\bar{x} = \frac{1}{|A|}\sum_{i \in A} x_i$ represents the average voxel intensity in the considered volume *A*. $|A|$ is the number of voxels in *A* [18]. Equation 2.3 means r is the ratio of the covariance of the voxel intensities contained within the volume of interest to the product of their standard deviation. r ranges between -1 and 1, where $r = -1$ means that the voxel intensities being compared are inversely proportional. When $r = 0$ there is no correlation between the voxel groups. If $r = 1$ then the voxel intensities in the two groups are directly proportional and there is a perfect match.

Although both DSI and *r* are used in this study, they are not directly comparable because the former describes the similarity of volumes while the latter describes the similarity of voxel intensities. However, this difference makes the DSI and *r* complementary to each other, so that when used together there is no redundancy in the information they provide [18].

3. Methods & Materials

The investigations carried out in this study to assess the accuracy of two Deformable Image Registration (DIR) algorithms are divided into 2 categories: phantom studies and clinical studies.

The phantom studies were designed to investigate DIR from first principles in a controlled environment where anatomical changes can be simulated to a known characterisation in order to determine the accuracy of the deformable registration algorithm. In the context of radiotherapy, a phantom is an object that is designed to mimic human tissue. Water is the standard phantom material for dosimetry; however solid materials can be used to imitate water in 3 parameters: mass density, electron density, or effective atomic number [43]. In contrast to phantom studies, anatomical changes in studies using patient image sets cannot be controlled, and not all changes can be easily identified. Nonetheless, retrospective studies on clinical cases can still be assessed qualitatively, and quantitatively using analytical software tools and by Radiation Oncologist (RO) review.

All studies were performed using either a dedicated MATLAB workstation for DIRART or the test environment of the Eclipse treatment planning system (TPS) for SmartAdapt. Clinical patient images were exported from the clinical TPS to the test-system and anonymised before being processed to maintain confidentiality, as well as preventing modification of the clinical record.

3.1. Deformable Image Registration Protocol

A standardised process has been used to carry out DIR to ensure consistent results:

1. Rigid registration of the planning CT as the source image and the desired target image was performed using the bony anatomy in the patient or the reference structure in the phantom as basis of registration.

    a. The two images were superimposed approximately using manual alignment.

    b. A region of interest (ROI) that covers the common volume of both source and target images was placed on the images and automatic registration (auto-match) was performed using the tissue density range (200-1700 HU) for bony anatomy, as performed clinically.

    c. The result of the registration was qualitatively assessed using the 'image difference' or 'colour blend' viewing tools. For the clinical cases, the results of the auto-match were manually adjusted to achieve the best match in which the central vertebrae (C2 – C4) are given priority over other regions of the images. This is because this part of the spinal column is what is used in clinical registrations as it is easily identifiable in planar images, cone-beam CTs and planning CTs.

2. Automatic deformable image registration was carried out.

a. Before beginning DIR the ROI was adjusted to include only the common areas between the images, so that if one image is larger than the other, the ROI boundaries will stop at the edges of the smaller image. The default DIR algorithm was used to carry out the registrations (no smoothing).

b. After DIR, the deformation grid and magnitude colour map were inspected to assess the quality of the registration. The results were also examined to check whether the observed changes correlate to the perceivable differences between the images.

c. The contours of the volumes of interest (VOI) (Section 3.1.1) from the deformed source image were copied onto the target image.

3. The clinical registration results were referred to the RO, while the phantom results were reviewed by the experimenter. The RO used the contouring tools (brush and eraser) in the TPS to correct any deformed VOIs that the DIR algorithm did not accurately process.

4. For the clinical cases, the deformed VOIs from the source image were again copied onto the target image to result in 2 VOI sets being superimposed on the image: (I) the VOIs deformed by the DIR algorithm without modification, and (II) the deformed VOIs amended by the RO. Thus, the VOIs deformed by the DIR algorithm were compared with VOIs accepted by the RO. For the phantom studies, the experimenter delineated the VOIs onto the target image for comparison with the VOIs deformed by the DIR algorithm. Therefore, the VOIs amended and approved by the RO and those

drawn by the experimenter directly onto the target image were now considered as the Reference VOIs.

5. Since no similarity measures were explicitly available in the clinical DIR programme, the 'Boolean Operations' tool was used to calculate the volume of the intersection between the automatically deformed VOIs and the Reference VOIs. Using this information and the original VOI volume statistics, it is now possible to calculate the Dice Similarity Index (DSI) as per Equation (2.2) in Chapter 2.

3.1.1. Selection of Structures for Analysis

Typically, there are many structures delineated on a given planning CT (pCT) that span physical volumes such as organs at risk (OAR) and gross tumour volumes (GTV), or geometrical volumes such as the planning target volumes (PTVs), which corresponds to margin added around a physical volume like the clinical target volume (CTV). OARs and GTVs delineate patient anatomy to ensure the delivered dose is within tolerance. Dose information for these structures is also used in plan optimisation and comparison. All these structures are contained within the 'Body' structure that encompasses the whole scanned volume of the patient anatomy as well as any immobilisation devices.

The structures selected for analysis in this investigation were those deemed of greatest clinical significance: GTV, CTV(s), PTV(s), spinal cord, parotid glands, and the brainstem. In radiotherapy planning, the PTV is the VOI to which treatment is prescribed and its adequate coverage is one of the main goals of the treatment, as it is a volume that contains the GTV and CTV to ensure they receive the prescribed dose [3]. In addition, the C3 vertebra

was contoured to act as a reference structure not expected to deform, and was useful when considering contouring uncertainty (Section 4.1.2).

3.2. DIR Programmes

3.2.1. DIRART

DIRART is an open source software tool which can be downloaded from the internet [44]. It has been developed using MATLAB, and is a research-purposed, non-clinical programme designed to perform adaptive radiotherapy (ART) functions via DIR algorithms, visualisation features and dose metrics analysis functions [34]. The version of DIRART that was available for this study was designed to carry out automatic DIR on image matches of the same modality, i.e. CT-CT, CBCT-CBCT, or MRI-MRI. DIRART has two main functional components (Figure 3.1), one for image registration and one for ART [25]. These components are linked by voxel mapping that tracks the observed displacement of each voxel in the source image to its corresponding location in the target image. For the purposes of this study, the performances of 4 of DIRART's collection of DIR algorithms were assessed.

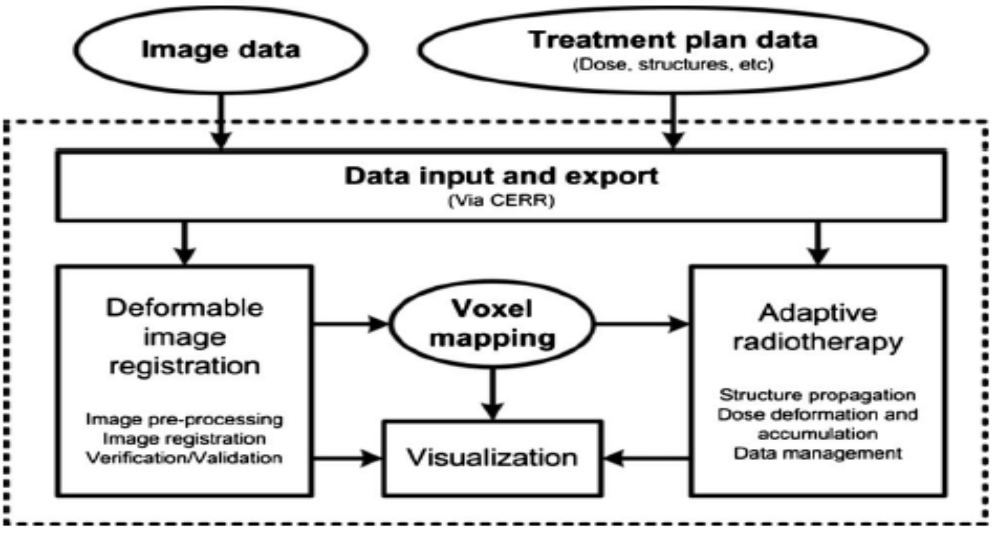

Figure 3.1: Schematic organisation of DIRART where blocks are functional components and ovals are data. Arrows are data flow, dotted line encloses DIRART components [34].

Image Registration

Rigid image registration in DIRART only allows translations in the 3 cardinal directions (axial, sagittal and coronal) without rotations or scaling[34]. Deformable registration is in turn implemented in 2 frameworks: the first is the default framework where the source image is deformed to the target image. This framework is referred to as the 'asymmetric' framework since one image is deformed while the other is kept fixed. The second framework is called a 'symmetric' or inverse consistency registration framework because both images are deformed towards each other [34].

Validation

A number of similarity measures (SM) were included in DIRART. Of the four similarity measures implemented, the Pearson's product moment correlation coefficient (*r*) was selected for use in this study, as its scoring is similar to the scoring of the Dice Similarity Index that was available in DIRART.

DIRART also provided visualisation tools (arrows and mesh grids superimposed onto the registered images) to qualitatively assess the deformation vector field (DVF) described in Section 2.1 [34].

3.2.2. SmartAdapt

SmartAdapt® (SA) v10 is part of the Eclipse™ v10 Treatment Planning System (TPS) from Varian Medical Systems®. It is designed to perform rigid and deformable registration operations, contouring and structure manipulations, and image and structure analysis functions in a clinical environment, as used in the protocol described in Section 3.1 [31]. It is designed to carry out automatic DIR on CT-CT, CT-CBCT, and CBCT-CBCT image pairs,

by employing a deformation algorithm based on the Modified Demons algorithm (Section 2.2.3) [32].

Rigid Image Registration

Like DIRART, the rigid registration component of SA allows for translational motion of the source image in relation to the target image, but it also allows for rotational motion including roll, pitch and yaw [31]. The Hounsfield Unit (HU) filter was used to specify an HU range of 200-1700 HU so that bony anatomy is used to match the images.

Deformable Image Registration

In SmartAdapt only the asymmetric DIR framework is implemented and only the default Modified Demons algorithm was used in this study. SA also has a variant of the default algorithm called the 'Smooth Algorithm', which has *a-priori* assumptions about the type of deformations that can be expected if the user specifies the anatomic region being registered. It was not used in this study because it can introduce over-corrections in the registration [31]. Moreover, smoothing would introduce a bias against the phantom studies when compared to the clinical studies, as the former did not always include an anatomic region. Therefore, only the default Demons algorithm without additional smoothing was used.

Figure 3.2 illustrates how the 2 images being registered are processed by the registration framework to implement the DIR algorithm. Essentially, the images are compared in the Similarity module to acquire a similarity value that is increased by the Optimiser to produce transform parameters. The transform parameters are used in the Interpolator to modify the voxels of the source image. The modified source image is then fed back into the Similarity module and the process is repeated until the similarity value is maximised. In this project, the default parameters for the Similarity, Interpolator and Optimiser modules were used for all registrations without any customisation.

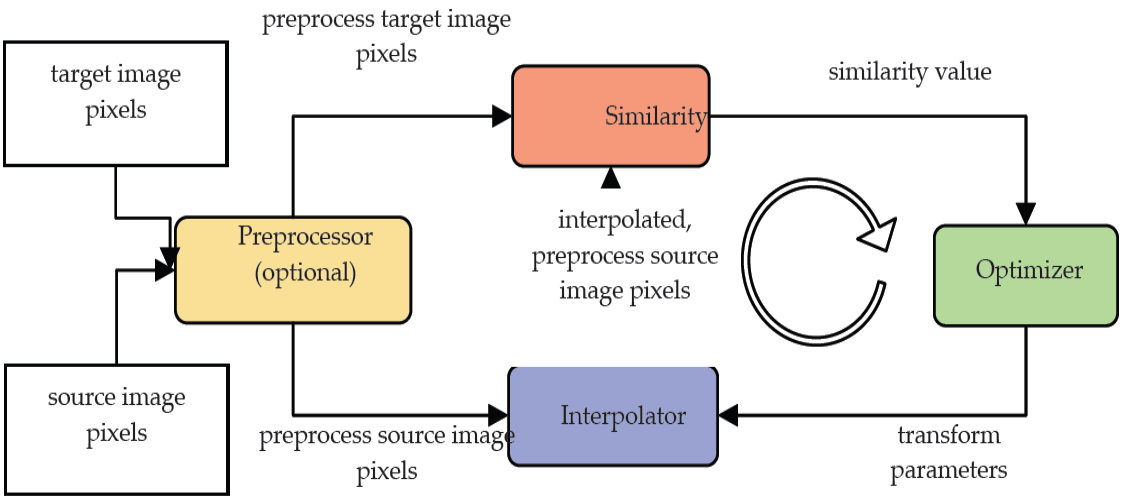

Figure 3.2: Schematic diagram of the image registration processes in SmartAdapt given in the SmartAdapt reference guide from Varian Medical Systems [31].

Validation

Image blending and image difference tools that superimpose the source and target images were used to qualitatively compare how closely the images matched each other after rigid registration and before DIR to how they matched after DIR. Colour-intensity maps were used for the visualisation of the perceived changes in the clinical patient images in order to reconcile them to changes noted in the patient's clinical record (Section 5.4.2). Quantitative analysis tools available in SA included determination of the volume change in the VOIs that was used in the DSI calculations.

3.3. Phantom Studies

The phantom studies were designed to separately study simulated rigid and non-rigid changes as well as the influence of CT slice thickness on overall DIR uncertainty (Figure 3.3).

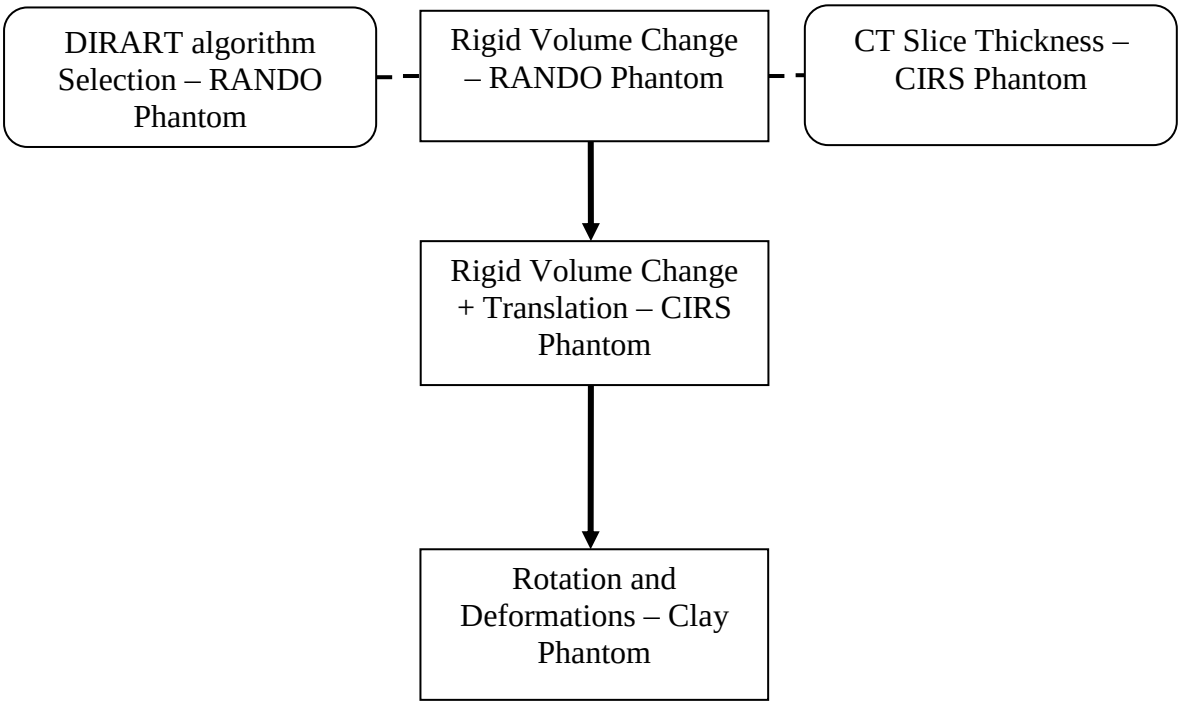

Figure 3.3: Outline of phantom investigations designed for this study. Rectangular elements represent central investigations; rounded elements represent initial investigations. Solid arrow connectors represent the escalatory pattern of the investigations moving from simple rigid changes to deformations.

3.3.1. Initial Investigations

Two investigations were designed to establish: A) the most suitable and accurate DIR algorithm in DIRART to apply to the primary phantom investigations. B) the effect of CT slice-thickness on DIR algorithm accuracy and performance.

DIRART Algorithm Selection

The varying developmental status of the DIRART code for the different algorithms available therein made it necessary to design an experiment to assess the suitability and accuracy of these algorithms. The DIRART User Manual provided indications on the functionality of most algorithms, so that a pool of 6 algorithms deemed to be 'working' was summarised. Of those algorithms, 4 were selected for testing based on their principal method, i.e. 2 based on the Optical Flow method, 1based on the Free-Form via Calculus of Variance method, and 1from the Demons based method. This initial selection was chosen to provide as diverse a range of methods as achievable for comparison. The remaining 2 algorithms were not tested as they were variants of the other 4 algorithms with minor additions or changes (e.g. the addition of a divergence constraint) that would not have provided significant added value to the investigation.

The algorithm selection for DIRART was carried out using the images obtained with the RANDO breast phantom as described below in section 3.3.2. For that purpose, the acquired images were registered according to the protocol outlined in Section 3.1 with the selected 4 algorithm variants: (I) Horn-Schunck (HS), (II) combined Lucas-Kanade and Horn-Schunck (HS+LK), (III) Optical Free Flow (FF), and (IV) Fast Demons with elastic regularisation (D +ER).

The selected algorithms were assessed and compared on two aspects: (I) qualitative accuracy of the DIR including the distribution of the DVF and its correlation with observed and expected change in the VOI, and (II) quantitative assessment of algorithm performance based on the *r* similarity measure scores (Equation 2.3 in Chapter 2).

CT Slice Resolution Uncertainty

This part of the study was designed to assess the impact of CT slice thickness on contour volume delineation in the planning system, which in turn can affect the ability of the DIR algorithm to determine the contour boundaries. CT slice thickness is known to affect an image's low-contrast resolution as well as its susceptibility to image artefacts due to partial volume effects, and blurring due to noise [45]. Moreover, slice-ambiguity can arise from the arbitrary selection of the initial start position of the CT scan; that is, the edges of the actual VOI may fall at the beginning or middle of the first slice. Due to the inherent averaging of tissue density between slices, the most distant parts of the VOI may either be just included or excluded in a specific CT slice depending on the start position of the CT scan, affecting the observed magnitude of the VOI. This investigation was performed using the CIRS phantom as described in section 3.3.3.

The phantom was scanned in 2 configurations comprising a VOI of 5 (configuration I) or 3 (configuration II) bone-tissue equivalent disc inserts (Figure 3.4). CT scans of both configurations were registered, as per protocol (Section 3.1), to simulate an enlargement of the VOI by choosing configuration I as the source image configuration II as the target image.

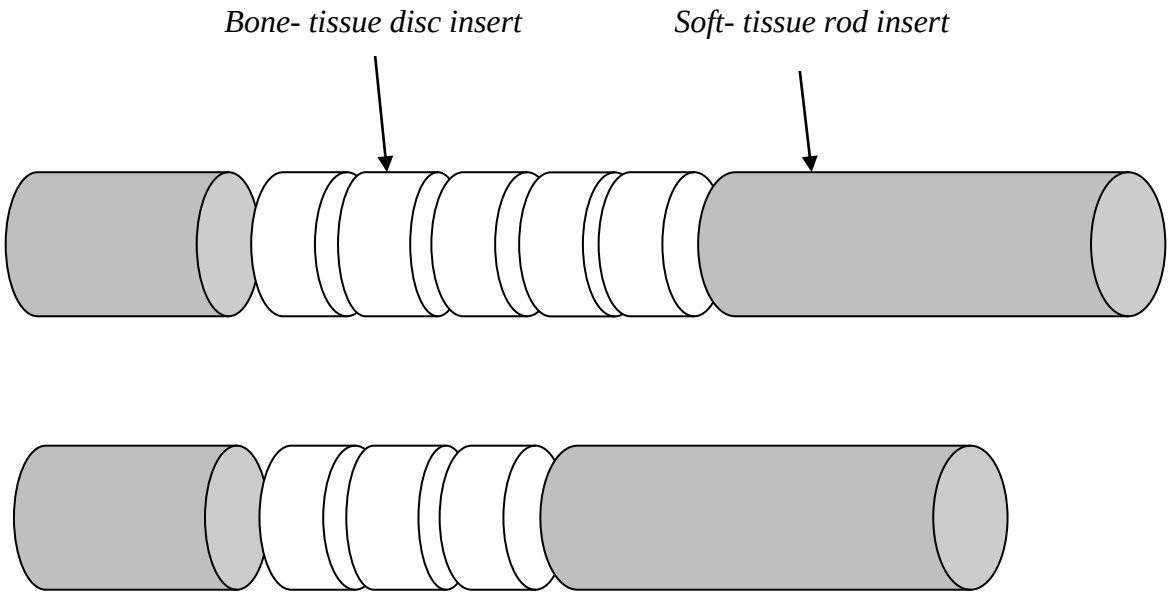

Figure 3.4: Configuration I of the CIRS phantom inserts for CT resolution study (top) comprising a VOI of 5 bone-equivalent discs, and configuration II with 3 bone-equivalent disc inserts making up the VOI (bottom).

The CIRS® phantom was CT-scanned in both configurations at various longitudinal resolutions with slice thicknesses in the z-direction ranging from 1 to 5 mm as indicated in Table 3.1. Actual volume changes were determined using calliper measurements, and compared to the volume changes observed after delineating the VOIs using the auto-contour tools in Eclipse.

Table 3.1: CIRS phantom scan matrix for the CT slice thickness uncertainty investigation. Each ● represents the two configurations illustrated in Figure 3.4. Scan position 'A + x mm' indicates the arbitrary scan starting position plus the offset introduced to accommodate slice-ambiguity

Scan Position	**CT Slice Thickness**			
	1 mm	**2 mm**	**3 mm**	**5 mm**
A	●	●	●	●
A + 0.5 mm	●	●		
A + 1.0 mm	●	●	●	●
A + 1.5 mm		●		
A + 2.0 mm			●	●
A + 3.0 mm				●
A + 4.0 mm				●

3.3.2. Volume Change Investigation – RANDO® Phantom

This investigation used the image sets of the gross volume change simulated in the RANDO phantom (Figure 3.5). The RANDO® phantom is an anthropomorphic phantom constructed with a natural human skeleton cast inside material that is radiologically equivalent to soft tissue. This material is made from a urethane formulation with an effective atomic number and mass density that closely simulates muscle tissue with randomly distributed fat [46]. The female version of the phantom used in this study was accompanied by breast attachments that are moulded into 2 cm thick sections and drilled on a 2 cm grid and held together with a

nylon screw. One of these breast attachments was used to simulate a gross volume change by removing a single section of the breast. Figure 3.5 shows the two configurations of the volume of interest (VOI) when the phantom was scanned. The whole breast was selected as VOI for DIR.

The phantom was first scanned in the CT scanner with 2 sections of a breast attachment fixed onto the phantom and the image was designated as the source image. After removing one breast section from the attachment the phantom was scanned again and the resulting image designated as the target image. The successive images were exported to Eclipse's SmartAdapt workspace and to the DIRART MATLAB-equipped workstation for processing.

Figure 3.5: Configuration of the RANDO phantom breast attachments showing the two applied configurations. A "shrinkage" or "growth" of the VOI can be simulated by swapping the respective images.

DIR was carried out as per the protocol in Section 3.1 in both SA and DIRART and the results assessed qualitatively by visual inspection of how well the deformed image matched the target image. In addition, the correlation of the DVF with the observed changes was also assessed using the colour map tool and 'show motion field' tool in SA and DIRART respectively. Quantitatively, the results were analysed by calculating the DSI and *r* for the VOI after DIR in SA and DIRART respectively, and by comparing the volume change observed by the DIR algorithm and the actual volume change induced.

This investigation has a VOI that is external to the main 'body' of the phantom, which means the induced change only occurs in the VOI and does not affect other parts of the phantom. This scenario was used as an example analogy of some clinical situations such as a large superficial tumour protruding from the patient's neck that shrinks during treatment and so the change only affects this VOI.

3.3.3. Volume Change + Translation Investigation – CIRS® Head & Neck Phantom

This investigation was carried out using the CIRS® phantom which is circular in shape and approximates the size of an adult patient's head and neck anatomy. It is constructed of soft-tissue equivalent epoxy materials (HU≈0) and is accompanied by various tissue equivalent interchangeable rod inserts and disc inserts (Figure 3.6) [47].

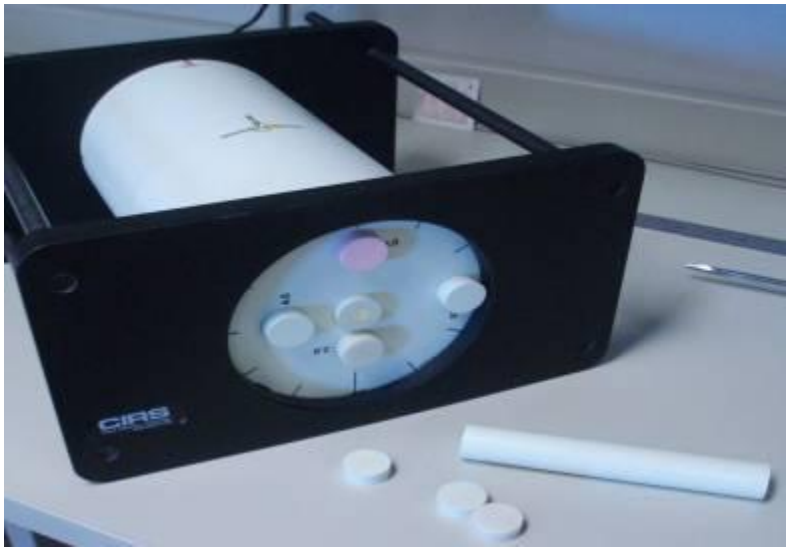

Figure 3.6: The head & neck CIRS phantom with some of its inserts. The small 1.1cm thick disc inserts have the same diameter of 2.53 cm as the rod inserts.

The investigation simulated a volume change with an added translation of the VOI by 1cm in the superior direction using the CIRS phantom and a number of its inserts customised for the purposes of the experiment (Figure 3.7).

The volume of interest was constructed of 3 soft-tissue equivalent discs that have been drilled to accommodate cavities filled with gypsum putty with 3 different diameters and a density that is ~250 HU in CT (Figure 3.7). The VOI was then sandwiched between soft-tissue equivalent rod inserts (Figure 3.8).

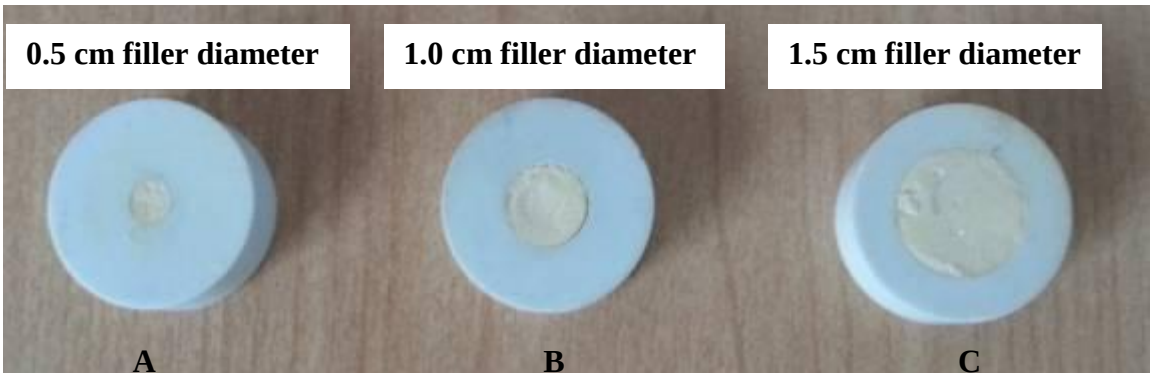

Figure 3.7 Modified CIRS phantom disc inserts containing the filler material that is delineated as the VOI on the CT scans of the phantom. Discs A, B, and C have internal diameters of 0.5, 1.0, and 1.5 cm respectively. All three discs have a thickness of 1cm.

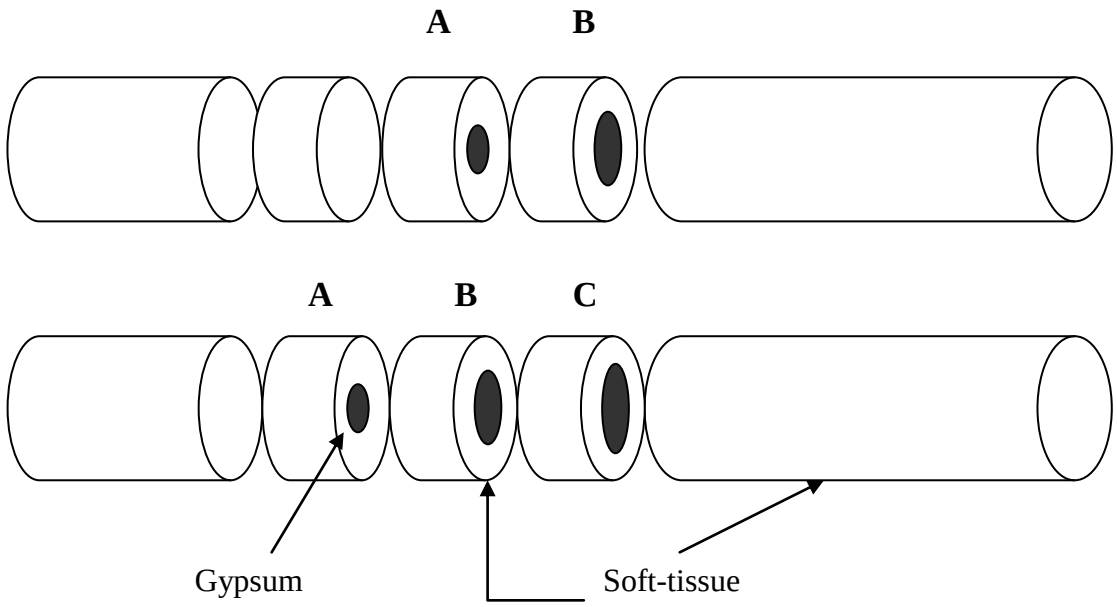

Figure 3.8: The initial configuration of the CIRS phantom inserts comprises the A and B putty-filled disc inserts (top). Volume change and translation was simulated by removing the unmodified soft-tissue equivalent disc insert preceding disc A while adding disc C after disc B and replacing the long rod insert (bottom diagram).

The phantom was secured to the CT table and scanned (this is the source image). To simulate a volume change and translation in this study, disc insert C (Figure 3.8) was added between the long rod and disc B, while the undrilled soft-tissue equivalent disc was removed. The phantom was scanned again to give the target image. For this investigation, DIR was carried out only in SA as it was not possible to use DIRART due to technical difficulties.

The same qualitative and quantitative assessments carried out for the investigation in Section 3.3.2 were done for this investigation, but only the SA tool (DVF colour map) and metric (DSI) were used in the analysis.

In contrast to the previous investigation, this experiment has a VOI that is inside the main 'body' of the phantom, which means that while the induced change affects the VOI, it also affects the 'anatomy' immediately surrounding the VOI. That is, if the VOI was enlarged and moved then the surrounding tissue was displaced to some extent to accommodate the change (removal of the unmodified soft-tissue disc). This scenario was used as an example analogy of clinical situations where a GTV situated deep into the patient anatomy may change in volume and shape in response to treatment, but it can also affect the location and shape of nearby OARs.

3.3.4. Rotation and Deformations – Clay Phantom Investigation

The purpose of this study was to simulate rigid rotation and deformations which were induced independently: (I) rotation, (II) bending, (III) pressing, and (IV) volume change. For this purpose, an in-house phantom was constructed from a modelling clay rod a fixed onto a solid water block. The design of the phantom had 2 main goals: a) the solid water acted as a reference rigid structure to the deformable clay rod to mimic the relationship between bony anatomy that is non-deformable and soft-tissue that is deformable. b) the solid water provides

HU/density contrast to the clay rod when scanned to make it easier to qualitatively assess the match between the deformed nad target image (See Figures 4.8, and 4.10-4.12 in Chapter 4). Figure 3.9 shows the Clay phantom in its 5 configurations.

Figure 3.9: The in-house Clay phantom in its 5 configurations: original, rotation, bending, pressing and cutting or volume change. The white micro-pore tape was used to secure paperclips bent into the shape of an arrow to delineate orientation in the CT scans.

The original configuration was scanned first and its image designated as the source image. Then, in each of the configurations illustrated in Figure 3.9 the phantom was scanned and the resultant images were designated as the target image in their respective registrations with the source image. DIR was carried out as per the protocol in Section 3.1 for each experiment. The same qualitative assessment as described in Section 3.3.3 was carried out for all the registrations. The quantitative assessment of all the registrations also involved comparison between observed and actual volume changes and DSI calculations, even though the rotation, bending, and pressing experiments did not have a volume change. For these three experiments, the DSI scoring and volume discrepancy between the deformed VOI and the target VOI were used to provide information about contouring uncertainty (see Section 4.4).

3.4. Clinical Studies

A cohort of 5 head and neck cancer (HNC) patients was selected for comprehensive analysis and clinical validation of DIR (Table 3.2). All patients selected were diagnosed with a primary disease site in the upper neck, such as the tonsils. In addition, all selected patients had multiple planning CTs (pCTs).

The purpose of this investigation was to provide information on the range of anatomical changes in the patient cohort. Before performing DIR on the pCT sets for each patient, the contours delineated by the RO on each scan were compared to determine in what way, if any, a particular structure's volume had changed. With reference to the patient's clinical record and consultation with the relevant RO, it was then possible to correlate any such changes with noted physical or physiological changes.

Table 3.2: Clinical information for the 5 HNC patients. L and R designate left and right, dose is prescribed to a standard PTV and an additional boost to sub-volume of the PTV.

	Patients				
Item	**P1**	**P2**	**P3**	**P4**	**P5**
1° Tumour Site	L Tonsil	L Tonsil	R Tonsil	Nasopharynx	L Tongue
Tumour Volume (cm^3)	14	89	26	83	20
Cancer Staging	T2N1M0	T3N2bM0	T1N0M0	T1N0M0	T2N2bM0
Number of CTs	2	3	2	2	2
Planning CT Date	13/9/12	11/10/12	24/10/12	15/8/12	28/9/12
IV Contrast	1^{st} CT only	1^{st} CT only	1^{st} CT only	1^{st} CT only	1^{st} CT only
Additional Imaging Modalities	-	PET	-	MRI	MRI
RT Type	Radical	Radical	Radical	Radical	Adjuvant
Prescribed Dose (Gy)	54, 66	54, 66	54, 66	54, 66	54, 60, 66
Fraction dose (Gy)	1.8, 2.2	1.8, 2.2	1.8, 2.2	1.8, 2.2	1.8, 2, 2.2
Surgery	Yes	No	No	No	Yes
Chemotherapy	Yes	Yes	No	Yes	No
RT Fraction 1 Delivery Date	1/10/2012	30/10/12	7/11/12	27/8/12	15/10/12
Weight Change between CTs	↓ 6.7 kg	↓ 6.1 kg CT1-CT2 ↑ 0.8 kg CT2-CT3	↓ 1.3 kg	↓ 7.4 kg	↓ 2.0 kg
Observed Changes	Tumour shrinkage	Patient position	Swelling	Patient position	Swelling

3.4.1. DIR Performance for CT-CT Matches

This investigation was used to analyse the DIR performance accuracy when compared to the volumes delineated by the RO both quantitatively and qualitatively. Deformable image registration was carried out as per protocol (Section 3.1) for CT-CT matches. Four patients had 2 pCTs each which gave 1 CT-CT match for each, and one patient had 3 planning CTs that resulted in 2 matches (CT1 to CT2 and CT2 to CT3) for this patient. DIR performance was assessed quantitatively by the DSI method of Similarity Measures and qualitatively by visual inspection of the DVF produced by DIR using the colour intensity map tool.

3.4.2. DIR Based on Local ROI for CT-CT Matches

This investigation was used to examine how the choice of the region of interest (ROI) affected the deformation results. Because DIR is an optimisation process that attempts to find the best solution for all the data in the ROI, the result of the deformation becomes a compromise determined by the in-built objectives and boundary conditions of the algorithm. Therefore, changing the ROI means altering the information that is input into the algorithm, and may influence how it behaves in deforming the source image.

The DIR matches in Section 3.4.1 using the ROI selection (Global ROI) specified in the registration protocol (Section 3.1) were repeated with a change to how the ROI is delineated. The new ROI, referred to as the local ROI, was drawn to only cover the parts of the image sets that contained the largest PTV. The image intensity range criterion was kept the same, i.e. 200-1700 HU range to cover bony anatomy. Due to the localisation of the disease sites of the 5 patients, the local ROI criteria still covered most of the spinal column, although some structures like the Brain Stem were now partially or wholly outside the ROI (Figure 3.10).

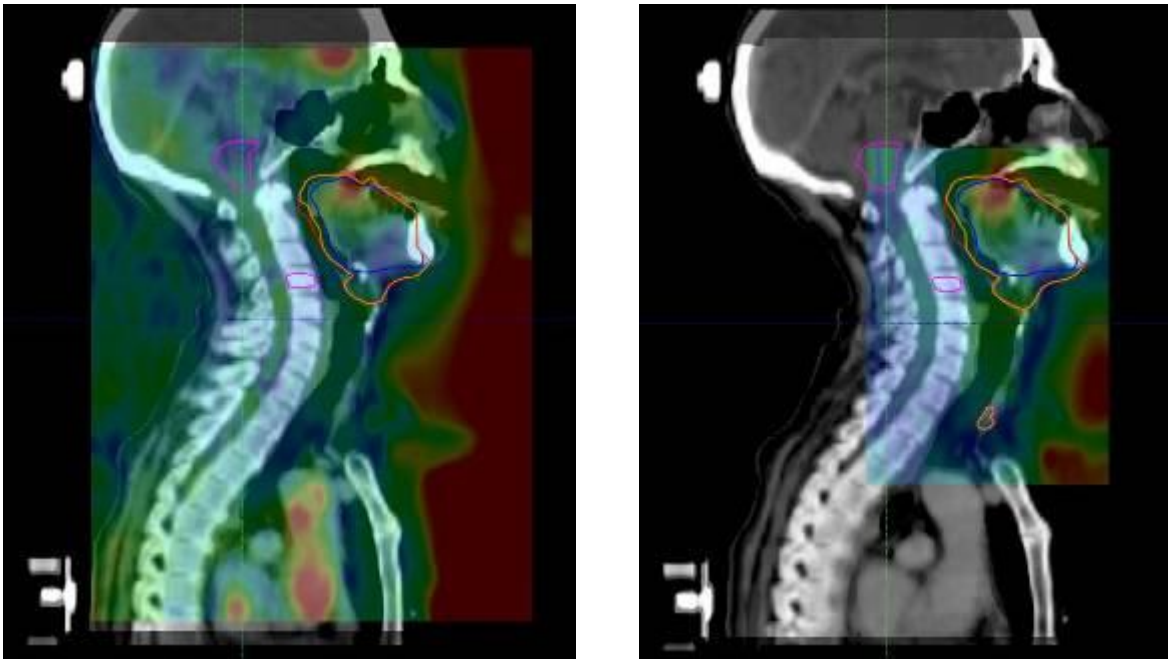

Figure 3.10: ROI boundary selection. The global ROI (left) as described in the registration protocol covers all of the common volume between the 2 registered images bar 1 increment from the edges of the smaller of the images. The local ROI (right) centres about the PTV and only covers its boundaries.

DSI scores measuring how the deformed VOIs matched those drawn by the RO were calculated for the registrations done with a local ROI. The average DSI scores across the patient cohort for all VOIs acquired with the local ROI registrations were compared to the average DSI scores obtained with the global ROI registrations.

4. Phantom Studies Results

The aim of these studies was to collect results from both DIRART and SmartAdapt. However, the use of DIRART met with multiple memory problems and functionality issues, so that the DIRART results were only acquired for those investigations using the RANDO phantom.

4.1. Initial Investigations Results

4.1.1. DIRART Algorithm Selection

RANDO phantom image sets simulating gross volume shrinkage were used in this study to select the type of algorithm from DIRART for use in the subsequent investigations. The four algorithms chosen for testing were original Horn-Schunck Optical Flow (HS), combined Horn-Schunck and Lucas-Kanade Optical Flow (HS+LK), Free-form via calculus of variations (FF), and Fast Demons with elastic regularisation (D+ER). The quality of image deformation and registration varied across the four algorithms (Figures 4.1 – 4.4), and this was reflected quantitatively in Table 4.1.

Figures 4.1 – 4.4 all show the same source and target images of the RANDO phantom registered together. The source image represents the phantom in configuration I where the VOI is the whole right breast, while the target image is a scan of the phantom in configuration II where one breast section was removed from the VOI. In the Figures' left panels, the source and target images are registered only rigidly, while in the right panels show the images after DIR where the source image is now deformed. The images are shown in 'difference' view, where the source image is rendered in a white colour palette and the target

image is rendered in a black colour palette, and areas of overlap between the images are gray as a result.

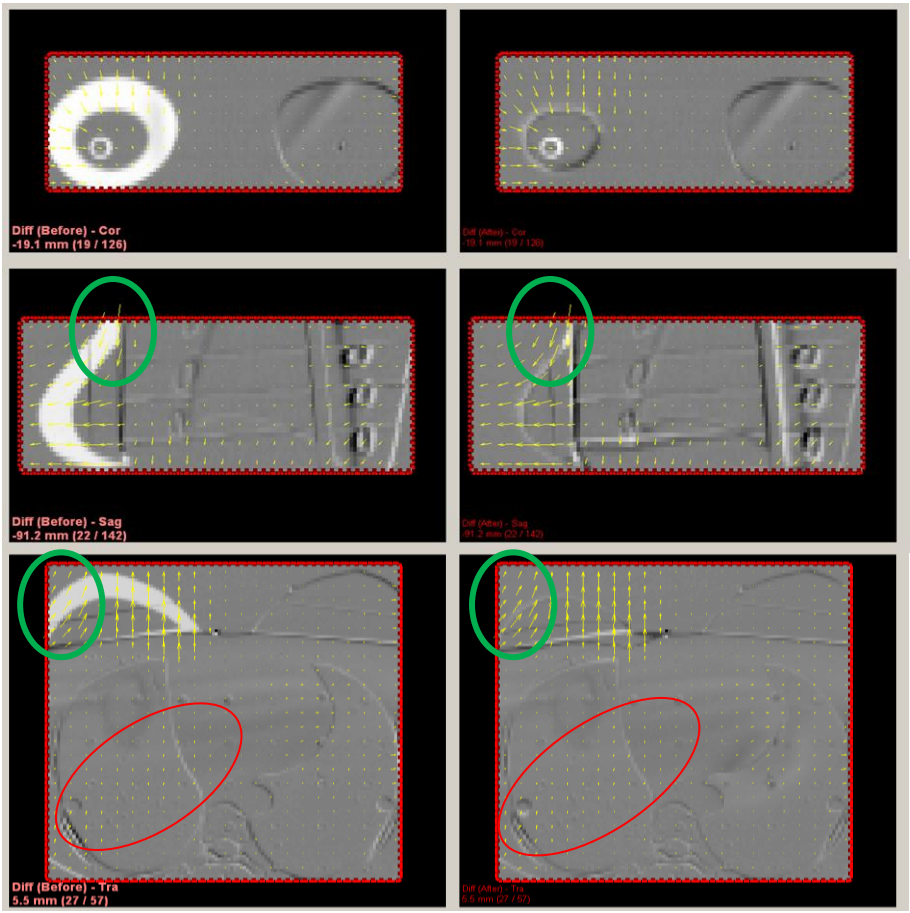

Figure 4.1: Superimposed images before (left) & after (right) DIR as produced by the Horn-Schunck Optical Flow Algorithm (HS). Yellow arrows overlaying the images represent the DVF computed by the algorithm. Green ellipses show erroneous deformation vectors not representing the correct direction of deformation. Red ellipses show non-zero vectors for region of the phantom where no change was induced.

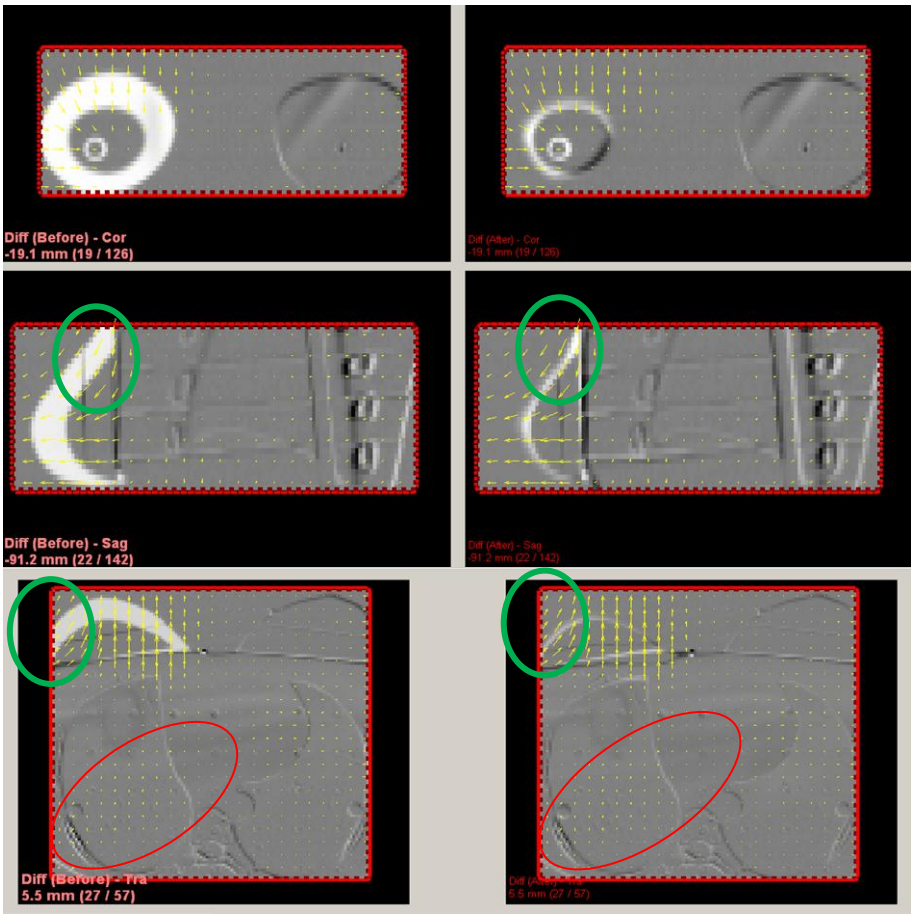

Figure 4.2: Superimposed images before (left) & after (right) DIR as produced by the Combined Horn-Schunck and Lucas-Kanade Optical Flow Algorithm (HS+LK). Yellow arrows overlaying the images represent the DVF computed by the algorithm. Green ellipses show erroneous deformation vectors not representing the correct direction of deformation. Red ellipses show non-zero vectors for region of the phantom where no change was induced.

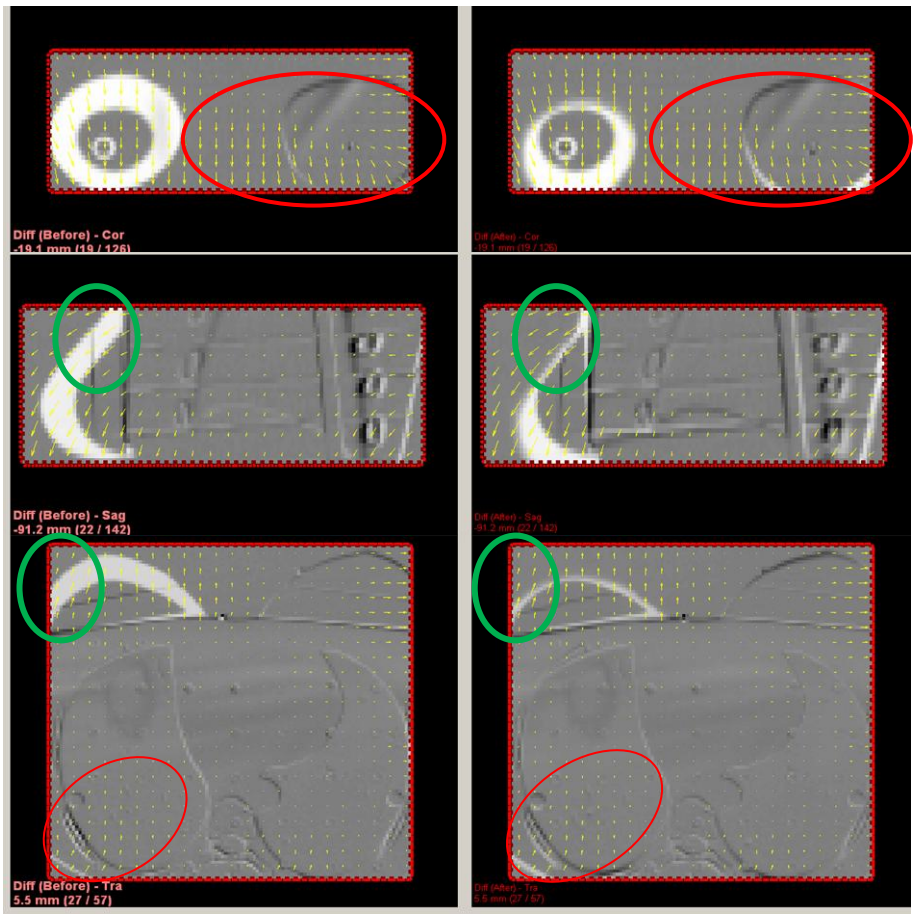

Figure 4.3: Superimposed images before (left) & after (right) DIR as produced by the Free Form Deformation Algorithm (FF). Yellow arrows overlaying the images represent the DVF computed by the algorithm. Green ellipses show erroneous deformation vectors not representing the correct direction of deformation. Red ellipses highlight the large false movement/change in the phantom interpreted by the algorithm. Red ellipses show non-zero vectors for region of the phantom where no change was induced.

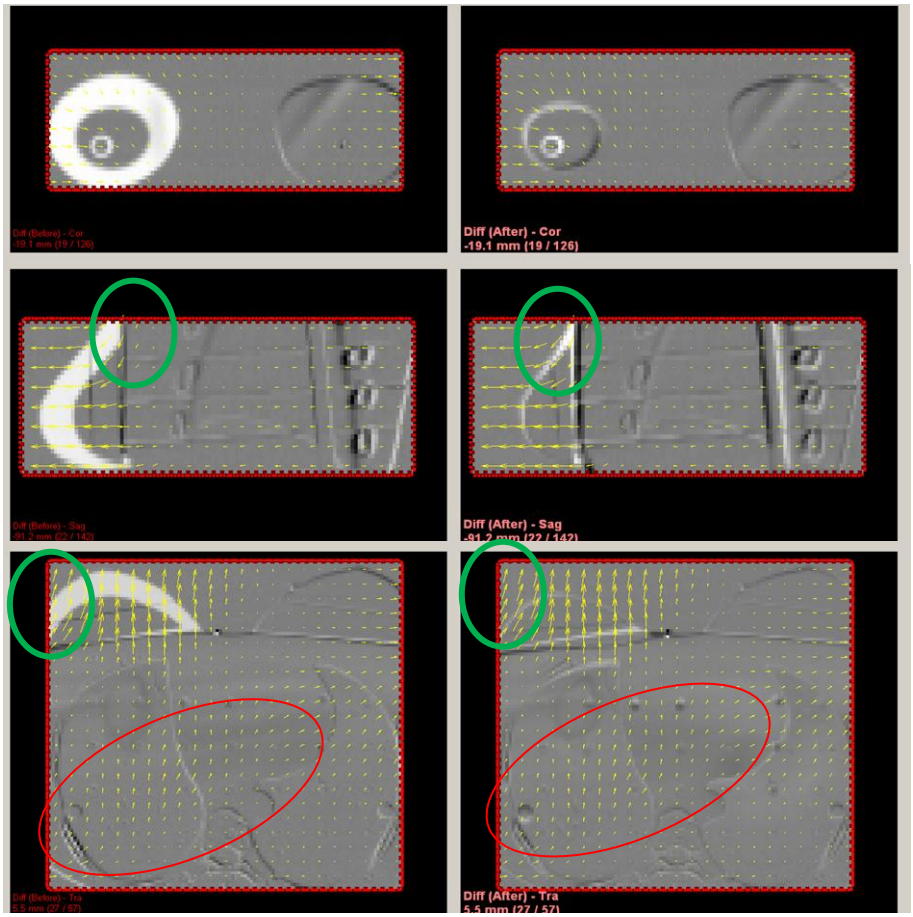

Figure 4.4: Rigidly registered images before (left) & after (right) DIR as produced by the Demons Algorithm with Elastic Regularisation (D + ER). Yellow arrows overlaying the images represent the DVF computed by the algorithm. Green ellipses show erroneous deformation vectors not representing the correct direction of deformation. Red ellipses show non-zero vectors for region of the phantom where no change was induced.

The actual volume of the removed breast section was 130 ± 1 cm^3, which was measured by determining the water displacement of the section using a small water tank and a measuring cylinder. The average volume difference noted by the treatment planning system (TPS) after 4 repetitions of manual contouring on both source and target images before registration was virtually equal and was 130.1 ± 1.5 cm^3. Automatic contouring of the VOI was not feasible due to the air gaps between the sections of the breast that would have confounded the Hounsfield Unit ranger used for auto-contouring. The size of the manually drawn VOI contours on the source and target images before DIR were 209.4 cm^3 and 81.8 cm^3, respectively. Quantitative analysis for all 4 algorithms was carried out using Pearson's product-moment correlation coefficient (*r*).

Table 4.1: Quantitative comparison of the performance of the 4 DIRART algorithms.

Algorithm	**Computation Speed (s)**	**Deformed Source VOI after DIR (cm^3)**	**Observed Source VOI Change after DIR (cm^3)**	**Difference between Target and deformed Source VOIs (cm^3)**	**Pearson's Correlation Coefficient, *r***
HS	341	101.6	-107.8	-19.8	0.85 ± 0.02
HS+LK	862	103.9	-105.5	-22.1	0.80 ± 0.03
FF	514	149.4	-60.0	-67.6	0.61 ± 0.06
D+ER	203	100.5	-108.9	-18.7	0.84 ±0.07

The HS Optical Flow algorithm gave the highest *r* score after deformation (0.85±0.02) of all algorithms. The *r* scores (0.84±0.07 and 0.80±0.03) obtained from the D+ER and the HS+LK algorithms respectively were similar to the HS algorithm, while the deformation results of the FF algorithm were poor (r =0.61±0.06) in comparison. Similarly, the HS, HS+LK, and

D+ER algorithms did better than the FF algorithm in terms of absolute volume. That is, the deformed source VOIs produced by the three former algorithms were closer in size to the target VOI (Table 4.1). In an ideal situation, the deformed source VOI should be exactly the same size as the target VOI, so the FF algorithm was demonstrably poorer in this regard. Overall, the quantitative analysis shows very similar results for the HS, HS+LK, and D+ER algorithms that perceptibly outperform the FF algorithm and thus narrow down the choices for selecting the most useable DIRART algorithm.

All the algorithms erroneously displayed non-zero deformation vectors in the areas of the image where the phantom was not deformed, such as over the breast contra-lateral to the VOI and over the lungs of the phantom (Figures 4.1-4.4). In particular, the FF algorithm gave the largest vectors of false movement (example circled red on Figure 4.3) while also showing how the deformed VOI did not match the target as seen by the white intensities around the greyscale target VOI. On the other hand, the 2 Optical Flow algorithms showed the smallest non-zero vector magnitudes in the unchanged regions, exhibited by the fewer and shorter vectors overlaying the areas of the phantom with no induced change. Further inspection of the DVFs of the HS, HS+LK and D+ER algorithms revealed there was a substantial difference in D+ER's perceived magnitude of the deformation. Namely, all the vectors computed by D+ER were systematically larger than the vectors computed by HS and HS + LK, which made the algorithm's DVF appear noisier, making it more difficult to discern the parts of the image showing actual change.

Furthermore, the DVF local to the VOI and proximal to the edges of the registration ROI (circled green on Figures 4.1-4.4) did not reflect the correct direction of voxel movement, i.e. the vectors in those regions reflected a sweeping or curved movement of the voxels while a linear, outward movement perpendicular to the chest wall was expected as indicated by the vectors on the other side of the VOI distal from the edges of the ROI. This expectation is

consistent with the direction of the volume change simulated by removing a section of the breast, i.e. the net result is the movement of voxel intensities towards the chest wall.

The HS algorithm gave the best match between the deformed and target images in comparison to the other algorithms, demonstrated by the smallest white intensity margin (representing the deformed VOI) around the target VOI (compare Figures 4.1-4.4). Essentially, if a deformation is completely accurate, i.e. the deformed source VOI exactly matches the target VOI, no white intensity would show. Therefore, the fewer white voxels, the more accurate the match. While the HS+LK algorithm and D+ER were of comparable qualitative deformation accuracy, the FF algorithm gave the poorest match where the source image was not deformed sufficiently and instead the VOI appears to have been shifted inferiorly rather than deformed (sagittal slice in Figure 4.3).

The FF algorithm showed the least favourable results both in qualitative and quantitative analysis. The other algorithms showed similar results in the quantitative analysis but the HS optical flow algorithm showed slightly better qualitative results and was therefore selected to be the most useable DIRART algorithm for the remainder of this study.

Furthermore, the HS algorithm had a relatively fast algorithm convergence speed (341 seconds), while the D+ER was the fastest in completing the registration in 203 seconds (Table 4.1). The HS+LK algorithm was the slowest (862 seconds), while the FF algorithm was also relatively slow (514 seconds). Algorithm processing speed was not an important factor in the phantom studies; however, it can have a significant impact on clinical workflow, if DIR is used in online adaptive radiotherapy for example.

4.1.2. CT Slice Resolution Investigation

To investigate the impact of CT-slice resolution and phantom scanning position, images of the CIRS phantom in two configurations (Section 3.3.1) were matched by DIR and quantitatively analysed by calculating the Dice Similarity Index (DSI) between the deformed VOI and the target VOI drawn on the target image.

Configuration I of the CIRS phantom had a VOI with an actual volume of 25.0 ± 0.0 cm^3 as measured using a Vernier calliper, while the actual volume of the VOI in configuration II was 15.0 ± 0.0 cm^3. The corresponding VOI generated in the TPS using auto-contouring varied for each source image and CT slice thickness (Table 4.2)

The difference between the volume of the deformed VOI and the target VOI displayed an oscillatory relationship with the scan position offset (Figure 4.5). This deviation does not appear to be directly related to slice thickness as the 3 mm slice thickness scans gave lower errors than the 1 mm slice thickness scans (Figure 4.5). The maximum error in volume shown in the graph was approximately 6 cm^3, obtained with the 5 mm and 2 mm resolution scans.

Table 4.2: DIR performance dependence on slice thickness for the CIRS phantom scanned in configurations I and II.

CT Slice Thickness (mm)	Source Image Contour Volume (cm^3)	Target Image Contour Volume (cm^3)	Mean Deformed Volume (cm^3)	Mean DSI
1	27.7 ± 0.1	16.6 ± 0.1	21.4 ± 0.3	0.82 ± 0.01
2	27.7 ± 0.0	16.6 ± 0.2	19.7 ± 2.0	0.83 ± 0.02
3	27.3 ± 0.0	16.6 ± 0.1	18.8 ± 0.6	0.81 ± 0.01
5	27.7 ± 0.0	16.6 ± 0.2	21.6 ± 1.7	0.82 ± 0.01

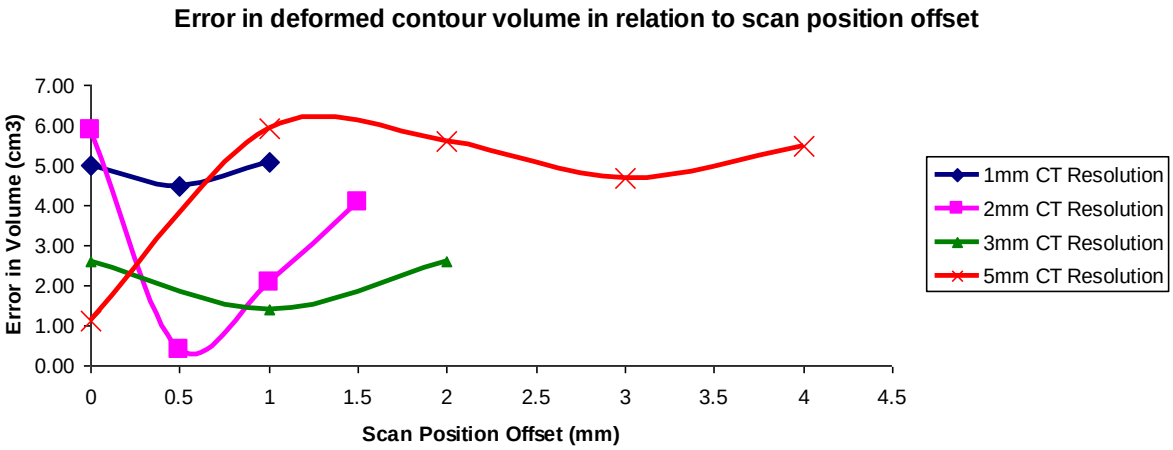

Figure 4.5: The oscillatory relationship between the scan position offset and the corresponding error in volume between the deformed contour and the automatically drawn contour in the TPS.

The results of this investigation showed that there were multiple sources of uncertainties that are inherent in the contribution of slice resolution to DIR algorithm performance, such as the auto-contouring Hounsfield Unit ranger tool in the TPS and the CT bed position during scanning. Therefore, analysing the results proved to be more complicated than anticipated previously because it was not possible to distinguish their particular contributions to the overall uncertainty. For example, the average deviations between the actual target VOI and the deformed VOI were 3.8 ± 0.6 cm^3 and 6.5 ± 0.3 cm^3 for the 3 mm and 1 mm images, respectively. In comparison, the average deviations between the auto-contoured target VOI and the deformed VOI were 2.2 ± 0.6 cm^3 and 4.9 ± 0.3 cm^3 for the 3 mm and 1 mm images respectively. This meant that the volume error related to slice resolution uncertainty was smaller than the overall error due to contouring, and so automatic contouring in the TPS was further dependent on other unknown sources of uncertainty. This experiment did not show a clear dependence of algorithm performance on CT slice thickness (Table 4.2).

4.2. Volume Change Results – RANDO Phantom

This investigation used the RANDO phantom (Figure 3.4) scanned in two configurations, as described in Section 3.3.2, to simulate a gross volume change. As stated in Section 4.1.1, the actual volume change simulated in the phantom was 130 ± 1 cm^3. Before DIR, the size of the manually drawn VOI contours on the source and target images were 209.7 ± 1.0 cm^3 and 79 ± 1.5 cm^3, respectively

Following DIR, the deformed source VOI was 94.0 ± 3.7 cm^3 in SA and 105.4 ± 2.9 cm^3 in DIRART (Table 4.3). These volumes were compared to the size of the VOI contour in the target image (79.6 ± 1.5 cm^3) to assess which algorithm gave the closest volume match.

Table 4.3: The VOI magnitudes for the four experiments carried out on the RANDO phantom images. The VOI contours manually drawn on the source and target images are given along with the difference between them. The deformed VOIs as determined by SmartAdapt and DIRART are presented for comparison with target image VOI.

Experiment	Source Image Contour VOI (cm^3)	Target Image Contour VOI (cm^3)	Volume Difference (cm^3)	Deformed Source VOI after DIR – SmartAdapt (cm^3)	Deformed Source VOI after DIR – DIRART (cm^3)
1	209.4	81.8	-127.6	91.3	104.09
2	209.8	78.6	-131.2	92.7	101.57
3	208.3	77.8	-130.5	91.5	106.29
4	211.2	80	-131.2	100.3	109.45
Mean ± 1 S.D.	209.7 ± 1.0	79.6 ± 1.5	-130.1 ± 1.5	94.0 ± 3.7	105.4 ± 2.9

Both algorithms (Modified Demons in SmartAdapt and HS in DIRART) gave similarity measure scores close to 1 (Table 4.4). SA gave an average DSI score of 0.90 ± 0.02, while the average *r* score obtained from DIRART was 0.85 ± 0.02.

Table 4.4: Quantitative comparison of SA and DIRART registration performance.

Exp	Target VOI (cm^3)	Smart Adapt			DIRART		
		Deformed VOI (cm^3)	Residual volume (cm^3)	DSI	Deformed VOI (cm^3)	Residual volume (cm^3)	*r*
1	81.8	91.3	9.5	0.92	104.1	22.3	0.85
2	78.6	92.7	14.1	0.90	101.6	23.0	0.83
3	77.8	91.5	13.7	0.89	106.3	28.5	0.83
4	80	100.3	20.3	0.87	109.5	29.5	0.89
Mean ± 1 S.D.	79.6 ± 1.5	94.0 ± 3.7	14.4 ± 3.9	0.90 ± 0.02	105.4 ± 2.9	25.8 ± 3.2	0.85 ± 0.02

The variability in the size of the same VOIs seen across the four experiments indicated that there was an uncertainty associated with the contouring process, which was relatively small when compared to the VOI changes. The relative uncertainties in contouring, calculated by dividing the standard deviation of the VOI by the VOI (Table 4.3), are 0.5% and 1.9% in the source image and target image, respectively.

To determine how the contour uncertainty is propagated in the DSI, which is a volume-based similarity measure, the relative uncertainties of the deformed VOI, target VOI, and the intersection volume were combined. The intersection volume, $|A \cap B|$ in equation 2.2, is the volume of overlap between the deformed VOI and the target VOI. This volume is independent from the source and target volumes. This is because it is created by the Boolean operations tool that constructs this contour using the intersecting boundaries of the two VOIs and not their absolute volumes. Therefore, combining the uncertainties of each volume (*A*, *B*,

and $|A \cap B|$) follows the 'law of the propagation of uncertainty', where the combined standard uncertainty u_c is equal to the positive square root of the estimated variance, as described in the NIST Reference on constants, units and uncertainty [48]. For Equation 2.2, where the three volumes form a quotient, the relative combined uncertainty $u_{c,r}$ formula given in the reference simplifies to $u_{c,r}(DSI) = \sqrt{{u_r}^2(x_1) + {u_r}^2(x_2) + {u_r}^2(x_3)}$, where x_1, x_2, x_3 are the standard deviations of *A*, *B*, and $|A \cap B|$ respectively. The average intersection volume was 77.7 ± 1.5 cm^3, which corresponds to a relative uncertainty of $u_r(|A \cap B|) = \frac{1.5}{77.7} \times 100 = 1.9\%$. Combining this uncertainty with the relative uncertainties of the deformed VOI and target VOI calculated in the same way (3.9% and 1.9%), gives the expected combined uncertainty in DSI, $u_{c,r}(DSI) = \sqrt{0.5^2 + 3.9^2 + 1.9^2} = 4.4\%$. In comparison, the actual relative uncertainty of the DSI, calculated from the estimated standard deviation in Table 4.4, was smaller (2.2%). It is important to explain why the calculated propagated uncertainty is larger than the estimated standard deviation. The propagated uncertainty takes into consideration the contouring uncertainty in the whole VOI. However, $|A \cap B|$ is only affected by errors in the volumes of *A* and *B* that it delineates, i.e. if *A* has a volume error in a region not overlapped by a region of *B* then $|A \cap B|$ is not involved and will not propagate this error. So in effect, the propagated uncertainty in the DSI is actually an upper limit that includes other sources of uncertainty, like CT slice thickness, that are not specific to the DSI uncertainty.

The qualitative results of the DIR matches in DIRART and SmartAdapt are compared in panels B and C of Figure 4.6. The left-hand side of the panels shows the source and target images after rigid registration and before DIR, while the right-hand side of the panels show the registration of the deformed source image to the target image after DIR. The yellow arrows in panel B represent the deformation vector field (DVF) produced by the HS

algorithm in DIRART. Similarly, the intensity colour map on the right-hand side of panel C represents the DVF calculated by the Modified Demons algorithm in SA.

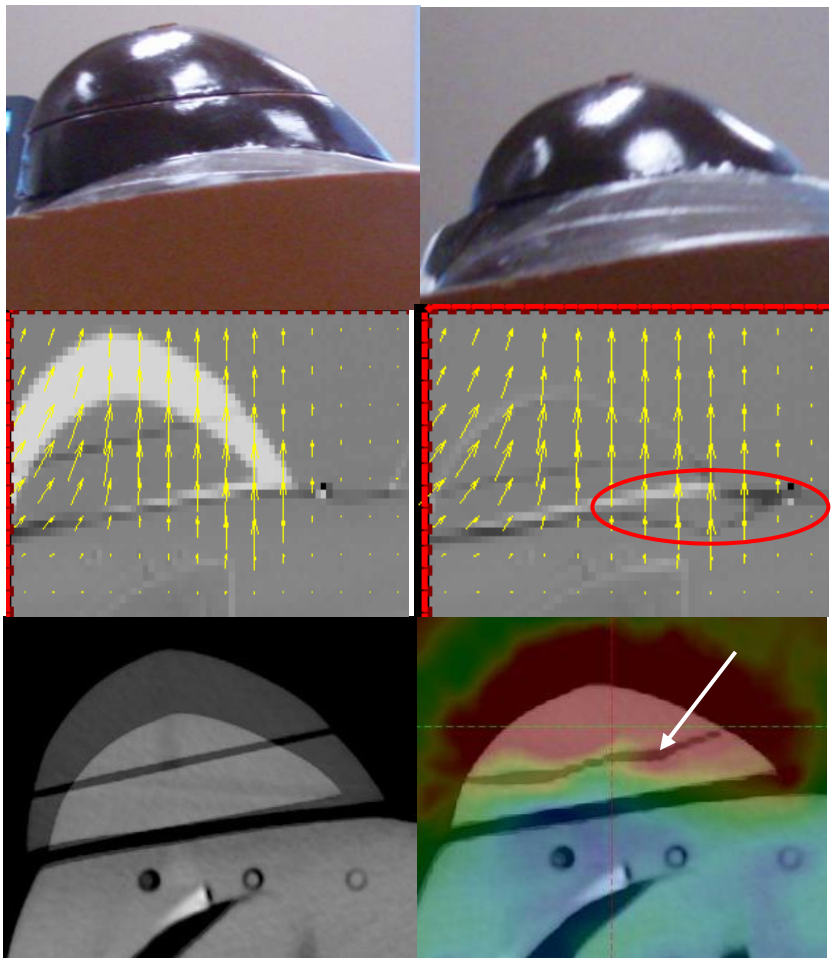

Figure 4.6: Qualitative example of the Modified Demons (SmartAdapt) and HS Optical Flow (DIRART) DIR results for the volume change in the RANDO phantom. A) Photographs of the 2 phantom configurations (Section 3.3.2). B) The registered images before and after DIR shown in DIRART; red ellipse shows encroachment of deformed VOI into main body of the phantom. C) The registered images before and after DIR shown in SmartAdapt.

Figure 4.6B illustrates how the HS algorithm in DIRART attempted to deform the source VOI to match the target VOI. As explained in Section 4.1.1, the registered images in DIRART are viewed in 'difference' mode where the source image is rendered in white intensity and the target image in black so that the areas of the images that coincide appear gray. From inspection of the figure, the deformed VOI matches the target VOI well, except at the chest wall interface and the anterior curve of the breast where a whitish margin can be seen. This margin corresponds to the residual volume from comparing the deformed VOI to

the target VOI summarised in Table 4.4. Additionally, the deformed VOI appears to partially extend into the chest wall (red ellipse on Figure 4.6B), which is a physical aberration that contributes to the larger size of the deformed VOI. The reason for this deformation error is difficult to determine; even assuming the air gap between the breast attachment and the chest wall confounds the algorithm's computations, the error would be expected to also affect the left side of the VOI, which it does not appear to do so.

Similarly, Figure 4.6C illustrates how the Modified Demons algorithm in SmartAdapt deformed the source VOI. In the colour map, reddish colours indicate high magnitude of deformation, greenish colours show moderate deformation, and bluish colours indicate little or no deformation. Inspection of the images after DIR (right) shows that the deformed VOI matches the target VOI very closely, giving a better result than the HS algorithm in DIRART, which agrees with the smaller residual volume resulting from SA for the comparison between source and target VOIs (Table 4.4). In spite of the better match, the Modified Demons algorithm produces its own error by warping the air gap between the two breast sections in the process of deforming the VOI (white arrow in Figure 4.6C). This example demonstrates one of the limitations of DIR algorithms that depend solely on voxel intensity distributions. The lack of the air gap in the target image confounds the algorithm's maximisation of the image similarity, which results in this physically unrealistic result where the algorithm can only force the air gap to fit the new volume.

4.3. Internal Volume Change – CIRS Phantom

In this investigation (Section 3.3.3, Figures 3.6-8) the VOI was located inside the body of the phantom where the first configuration comprised the source VOI made up of discs A and B

(see Figure 3.8) and the second configuration comprised all 3 putty-filled discs as the target VOI.

The calculated sizes of the source and target VOIs were respectively 1.0 cm^3 and 2.7 cm^3. The VOI was automatically contoured on the source and target images using the HU ranger. The contouring was repeated four times before carrying out DIR in SA (Table 4.5).

Table 4.5: Quantitative data for the CIRS volume change + translation study. The deformed VOI is compared to the target VOI across 4 experiments to obtain the DSI.

Experiment	**Source Image VOI (cm^3)**	**Target Image VOI (cm^3)**	**Deformed VOI (cm^3)**	**DSI**
1	1.2	2.9	1.5	0.66
2	1.3	2.8	1.5	0.69
3	1.2	2.8	1.6	0.70
4	1.3	2.9	1.6	0.70
Mean ± 1 S.D.	1.3 ± 0.1	2.8 ± 0.1	1.6± 0.1	0.69 ± 0.02

The average source and target image VOI contours were respectively 1.3 ± 0.1 cm^3 and 2.8 ± 0.1 cm^3, as reported by the TPS. After DIR, the average size of the deformed VOI copied onto the target image was 1.6 ± 0.1 cm^3. Quantitative analysis, calculating the DSI as described in Section 4.2, yielded an average DSI score of 0.69 ± 0.02 cm^3. To calculate the propagated uncertainty in DSI, the relative uncertainties for the target VOI and deformed VOI of 1.8% and 3.2% respectively were combined with the relative uncertainty in the intersection volume (3.4%), giving an upper limit of 5.0%. From Table 4.5, the actual relative uncertainty of the DSI came to 2.9% and hence within the calculated expected uncertainty.

The DSI scores obtained for this study were significantly lower than those in the RANDO phantom study (Section 4.2) and this is also reflected qualitatively in Figure 4.7. Figure 4.7d shows inaccurate deformation of the source image (circled in red), where the image is insufficiently deformed in the axial direction, i.e. the direction of volume change. In contrast, the yellow arrows point to how the algorithm deformed the VOI correctly in the lateral directions. One possible reason for this difference in performance is the presence of the air gaps between the disc-inserts making up VOI that can confound the algorithm; this seems to be a common limitation to these phantom studies that will be discussed in Chapter 6.

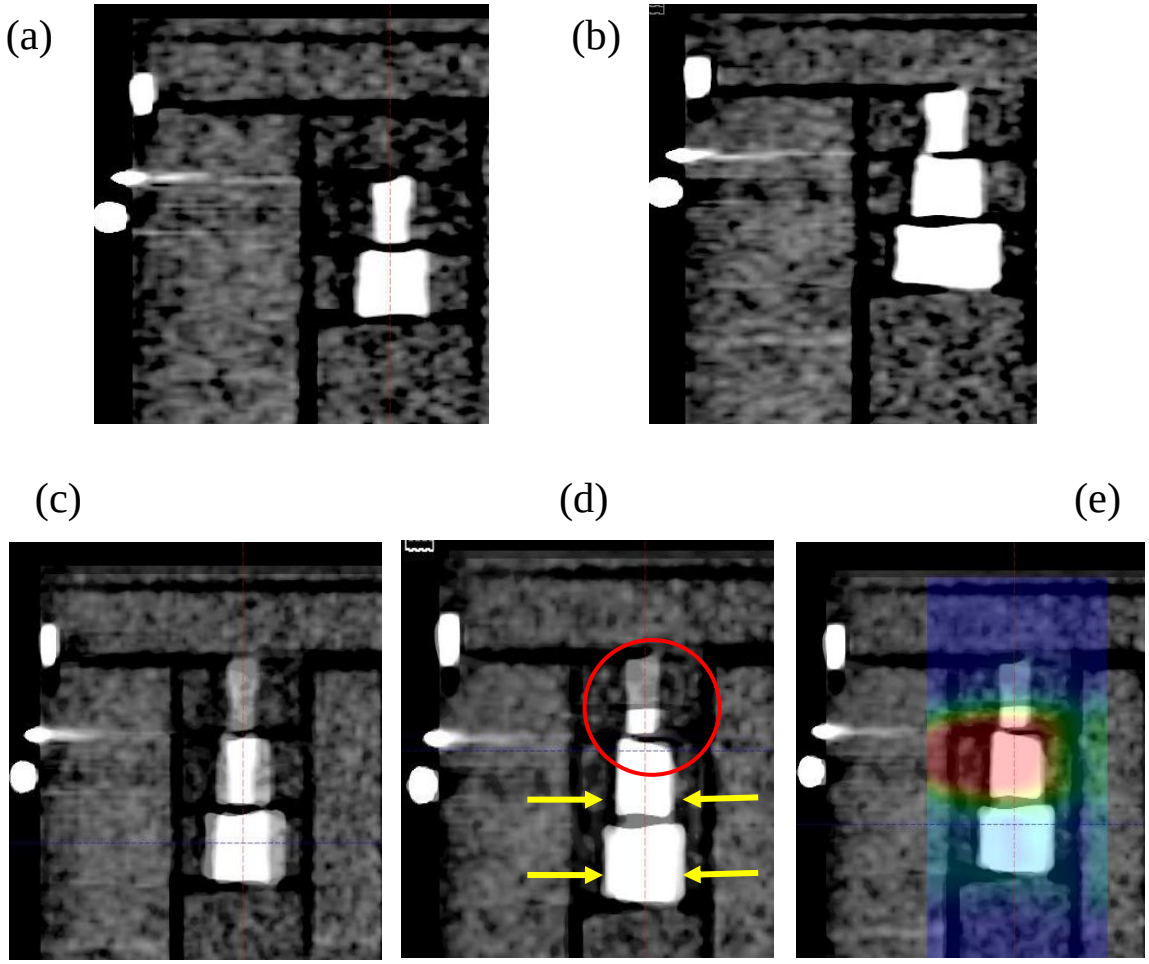

Figure 4.7: Coronal slices of the CIRS phantom through the centre of the VOI. (a) source image, (b) target image, (c) rigid registration of source and target before DIR, d) superposition of the deformed source image on the target image after DIR, e) superposition of the deformed source image on the target image after DIR overlaid with DVF magnitude intensity colour map.

4.4. Rotation and Deformations – Clay Phantom

In this investigation, a simple, in-house developed, deformable phantom was constructed from modelling clay and solid water resin, and used to simulate rotation and deformations of the VOI separately as described in Section 3.3.4. DIR was carried out using the Modified Demons algorithm in SA only.

4.4.1. Rotation

This experiment simulated a rotation of the VOI without deformation to test the boundary conditions of the Modified Demons algorithm when dealing with a relatively large rotation of ≈45º (Figure 4.8).

After the initial rigid registration using the solid water base as the registration target, the images in this experiment were first registered with DIR using a global ROI encompassing the whole phantom, including the solid water, as per the default action dictated by the DIR protocol (Section 3.1). The DIR produced a physically unrealistic result (Figure 4.8d). A local ROI was then used to cover only the rigidly registered clay rods, while excluding the solid water base. The results were greatly improved and gave a realistic deformation (Figure 4.8e).

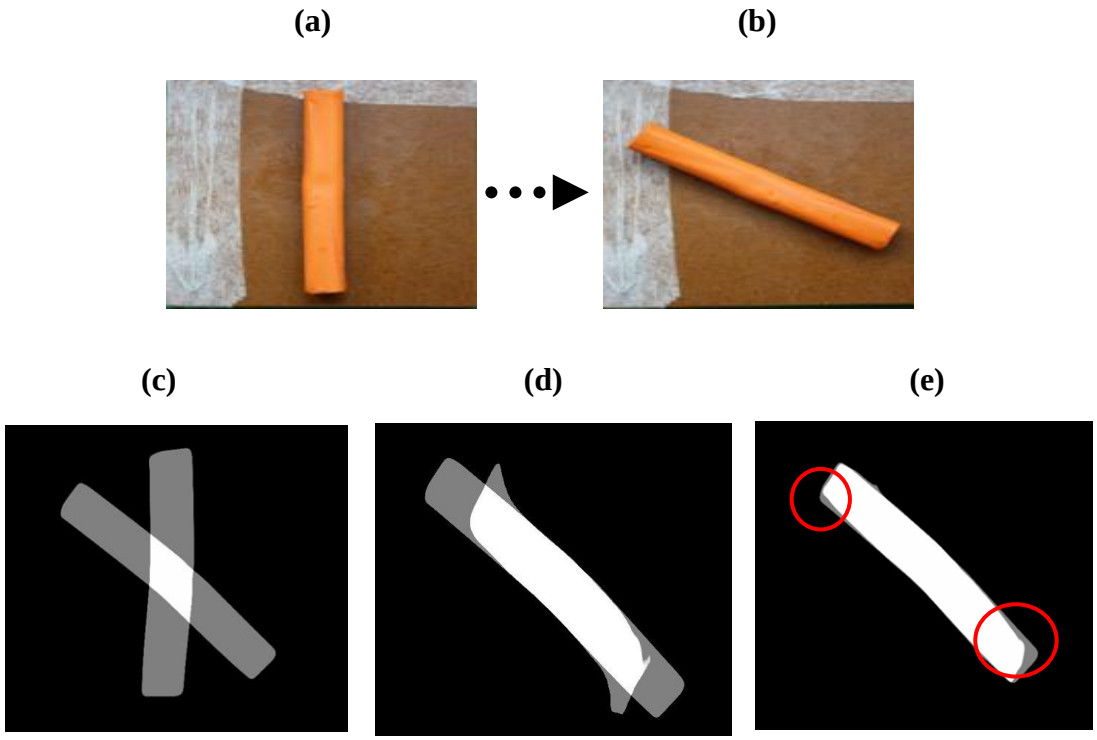

Figure 4.8: Photographs of the phantom in the original configuration and the rotation configuration (a and b). Central coronal image slices through the clay rod (bottom): c) source and target images are rigidly registered only d) images after DIR using a global ROI e) images after DIR using a local ROI encompassing only the volume of the superimposed rod images. Red circles show areas of mismatch.

The actual volume of the clay rod, measured with a ruler and Vernier callipers, was 7.4 cm^3. While automatic contouring of the VOIs on the source and target images gave an average volume of 7.3 ± 0.1 cm^3 after 4 repetitions.

Using the local ROI, the average DSI between target VOI and deformed VOI was 0.84 ± 0.02 cm^3 (Table 4.6). This table summarises the volumes needed for calculating the DSI, as defined in Equation 2.2, where the intersecting volume is related to the sum of the two volumes being compared (deformed source and target VOIs). The intersecting volume can also indirectly denote whether the VOIs being compared are displaced. For example, if the size of the deformed VOI is very similar to the target VOI but the intersecting volume is

relatively small, then it is possible that the deformed VOI is displaced from the target VOI, which can then be corroborated by qualitative analysis of the images.

Table 4.6: DIR results of the Clay phantom rotation experiment using a local ROI.

Experiments	**Target VOI (cm^3)**	**Deformed VOI (cm^3)**	**Intersecting Volume (cm^3)**	**DSI**
1	7.3	6.7	6.2	0.88
2	7.2	6.6	5.8	0.84
3	7.2	6.5	5.6	0.81
4	7.3	7.1	6.0	0.84
Mean ± 1 S.D.	7.3 ± 0.1	6.7 ± 0.2	5.9 ± 0.2	0.84 ± 0.02

Like in the RANDO and CIRS phantom investigations (Sections 4.2 and 4.3), a calculation of the propagated uncertainty in DSI was carried out for this experiment. Calculating the relative uncertainties for the target VOI and deformed VOI gave values of 1.3% and 2.9% respectively. Combining these results with the relative uncertainty in the intersection volume (3.4%) resulted in an expected relative upper limit of the DSI uncertainty of 4.7%. In contrast, the observed relative uncertainty in the DSI (2.4%) across the 4 trials was smaller than this calculated uncertainty.

4.4.2. Deformations

This section presents the results of the 'Bending' and 'Pressing' experiments of the Clay Phantom, termed 'Deformations'. The DIR results are presented qualitatively in Figure 4.9 alongside the rigidly registered images. Both DIRs were carried out using a global ROI, while a DIR done using a local ROI (not shown) gave identical results.

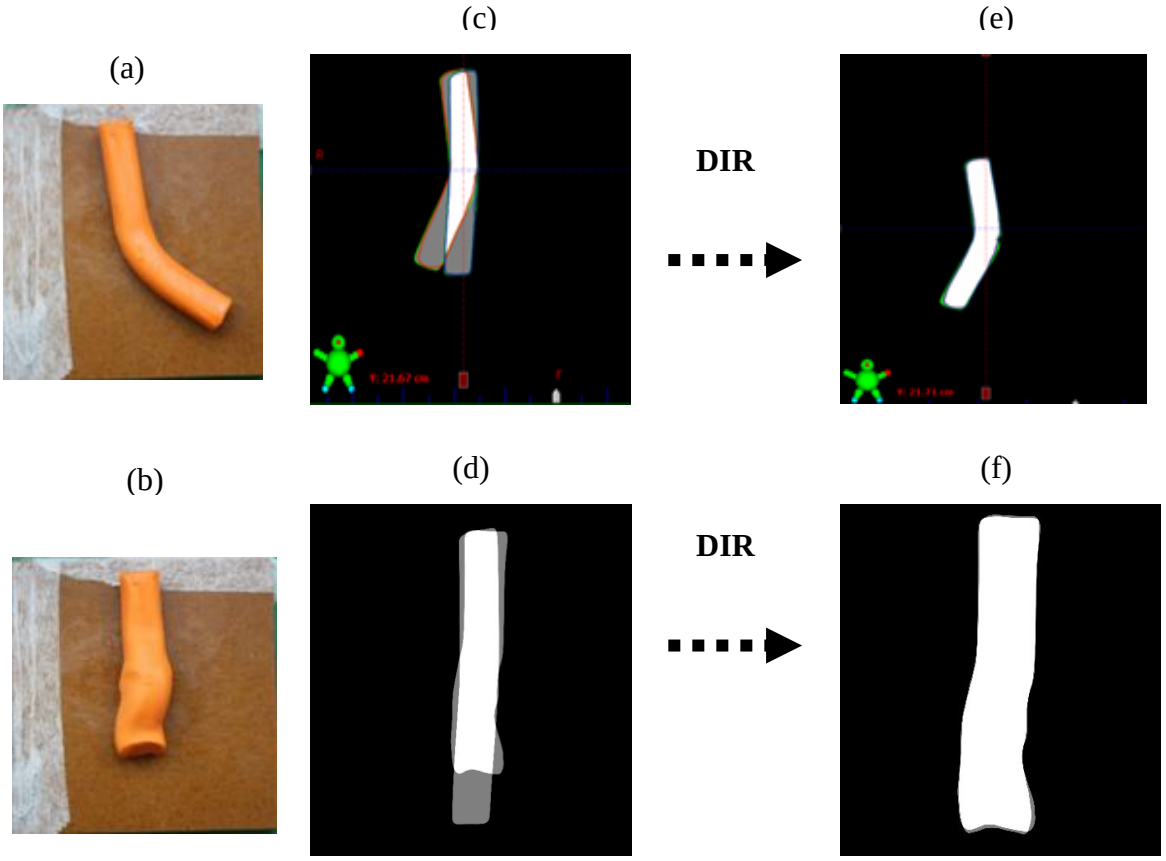

Figure 4.9: Qualitative example of the bending and pressing experiments, showing photographs of the two phantom configurations (a & b). Coronal slices of the target images rigidly registered to the source image of the original phantom configuration (c & d). Target and deformed source images after DIR (e & f).

The average DSI between target VOI and deformed VOI for the bending experiment was 0.92 ± 0.01 cm^3 (Table 4.7). The DSI score for the pressing experiment averaged at 0.93 ± 0.01 cm^3 (Table 4.8).

Table 4.7: Quantitative DIR results of the Clay phantom bending experiment from 4 repetitions.

Experiments	**Target VOI (cm^3)**	**Deformed VOI (cm^3)**	**Intersecting Volume (cm^3)**	**DSI**
1	7.3	6.8	6.5	0.92
2	7.2	7.2	6.8	0.94
3	7.2	6.9	6.5	0.92
4	7.3	6.7	6.3	0.91
Mean ± 1 S.D.	7.3 ± 0.1	6.9 ± 0.2	6.5 ± 0.2	0.92 ± 0.01

In the bending experiment, the relative contouring uncertainties of target and deformed VOIs were the same as those in the rotation experiment (1.3% and 2.9% respectively). Combining these results with the relative uncertainty in the intersection volume (3.1%) resulted in an expected relative uncertainty upper limit in the DSI of 4.4%. However, the observed relative uncertainty in the DSI across the 4 experiments was substantially smaller (1.1%).

Table 4.8: Quantitative DIR results of the Clay phantom pressing experiment from 4 repetitions.

Experiments	Target VOI (cm^3)	Deformed VOI (cm^3)	Intersecting Volume (cm^3)	DSI
1	7.3	7.3	6.8	0.93
2	7.2	7.2	6.6	0.92
3	7.3	7.0	6.6	0.92
4	7.2	7.1	6.7	0.94
Mean ± 1 S.D.	7.3 ± 0.1	7.2 ± 0.1	6.7 ± 0.1	0.93 ± 0.01

In the pressing experiment, the relative contouring uncertainties of target and deformed VOIs were 1.3% and 1.4%, respectively. Combining these results with the relative uncertainty in the intersection volume (1.5%) resulted in an expected relative uncertainty in the DSI of 2.4%. Again, the observed relative uncertainty in the DSI across the 4 experiments was smaller (1.1%). Overall, the bending and pressing experiments gave the most accurate DIR results qualitatively and quantitatively out of the phantom experiments. It is likely that this was due to the conservation of information in source and target images that does not require the DIR algorithm to account for missing grey values.

4.4.3. Irregular Volume Change (Cutting)

In this experiment, the Clay phantom was used to simulate irregular volume change by cutting a section from the rod to create deformed configuration (Figure 4.11). Due to the malleability and non-uniform shape of the removed portion, its volume was measured using water in a measuring cylinder and was equal to 3 ± 1 cm^3. The VOI was auto-contoured in the same method described for the other processes and the observed volume change was 2.8 ± 0.1 cm^3. The source and target images were registered in SA and analysed quantitatively (Table 4.9).

Table 4.9: Quantitative DIR results of the Clay phantom cutting experiment using a local ROI from 4 repetitions. *The standard deviations were actually 0.05, 0.04, and 0.13 cm^3 (2 s.f.) respectively.

Experiments	**Target VOI (cm^3)**	**Deformed VOI (cm^3)**	**Intersecting Volume (cm^3)**	**DSI**
1	4.5	4.4	4.1	0.92
2	4.5	4.4	3.9	0.88
3	4.4	4.4	4.2	0.95
4	4.4	4.3	3.9	0.90
Mean ± 1 S.D.	4.5 ± 0.1*	4.4 ± 0.0*	4.0 ± 0.1*	0.91 ± 0.03

A qualitative analysis of the DIR was carried out of the results for a global and local ROI. The global ROI resulted in a physically unrealistic deformation (Figure 4.10) similar to that observed in Section 4.4.1.

Figure 4.10: Sagittal slices of the DIR results in the cutting experiment using a global ROI (left) and a local ROI (right). Red rectangles delineate the ROI boundaries used in each situation.

Figure 4.10 also showed the DIR that used the local ROI also had a tapering in the deformed VOI in a similar position to the unrealistic deformation in the global ROI, but to a smaller extent.

The rigidly registered images in Figure 4.11c showed there was a small rotation in the VOI in addition to the induced volume change. This small rotation unintentionally added to the complexity of the simulated change in a minor way but did not appear to affect the results qualitatively.

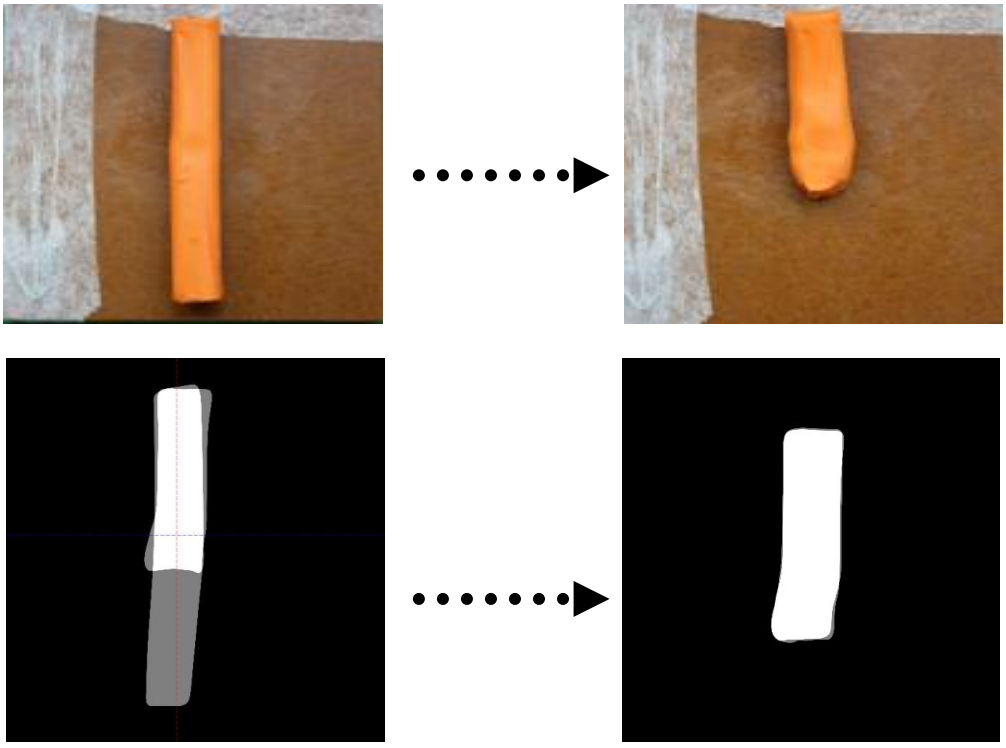

Figure 4.11: Top panel: Photographs of the 2 phantom configurations used in the cutting study. Bottom panel: Coronal slices of the source and target image after rigid registration only (left) and after DIR (right).

In this experiment the average DSI score comparing the deformed and target VOIs after registration was 0.91± 0.03 (Table 4.9). The relative contouring uncertainties of target and deformed VOIs were 1.1% and 0.9% respectively. Combining these results with the relative uncertainty in the intersection volume (3.3%) resulted in an expected relative uncertainty upper limit in the DSI of 3.6%. In comparison, the observed relative uncertainty in the DSI

across the 4 experiments was 3.3%. This experiment also gave relatively high DIR accuracy, even though it was a volume change involving missing grey values. However, unlike the RANDO and CIRS experiments, the VOI was of homogeneous intensity (no air gaps), which is less likely to confound the deformation computed by the algorithm.

5. Clinical Results

This chapter deals with DIR applied to images acquired during treatment of 5 patients as outlined in Section 3.4. The results for all 5 patients in the study were analysed quantitatively with SmartAdapt's statistical tools and qualitatively in consultation with the treating Radiation Oncologist (RO). The results are discussed considering the observed range of anatomical changes across the cohort, and the deformable image registration (DIR) performance of matches performed using a global region of interest (ROI) encompassing the whole image are compared to DIRs using a local ROI centred on the planning target volume (PTV). Sample images (Section 5.4) are used to summarise the qualitative results.

5.1. Observed Range of Anatomical Changes

An initial assessment of the range of anatomical modifications across the patients in this study was made by analysing the volume statistics of the structures delineated by the RO on each x-ray computed tomography (CT) scan (Table 5.1). In addition, the clinical notes in the patient records were used to identify the time scale between images as well as pertinent observed changes in the patient like weight change or tumour shrinkage.

In the following tables the target volumes (TVs) are the gross tumour volume (GTV), clinical target volume (CTV) and the PTV. R Parotid and L Parotid refer to the right and left parotid glands respectively. C3 indicates the third cervical vertebra contoured on each patient CT as a non-deformable structure. $t_{CT1\text{-}CT2}$ is the time between the original planning CT (pCT) and the second CT scans and $t_{F1\text{-}CT2}$ is the time between the delivery of the first fraction of radiotherapy to the time of the second CT scan.

Based on department protocol for head and neck patient contouring, the contouring of the TVs is carried out by the RO; the brain stem, spinal cord, and the parotid glands are considered critical structures for this disease site and are also contoured by the RO [49]. For all 5 patients, the initial planning CTs were scanned with intravenous contrast administered, while the additional CT scans were done without contrast. Intravenous contrast, containing iodine, is used to enhance the appearance of the larger blood vessels in the head and neck on CT scans so that they appear in a discernibly lighter grey than surrounding soft tissue. This can result in better defined boundaries for soft tissues such as the parotid or GTV if they border or are in close proximity to these blood vessels. This implies that contouring of VOIs in contrast-enhanced CTs will differ from contouring done in CTs with no contrast, because the RO may be able to distinguish tissues more clearly in the former. TVs and OARs may uptake miniscule amounts of IV contrast that cause voxel intensities of soft-tissues neighbouring the blood vessels to have slightly higher grey values than in the absence of contrast. The resulting difference in voxel intensity may be too small to be distinguished by the DIR algorithm, whereas the human brain processes grey scale images much better than a computer, so that an RO experienced in contouring can distinguish different soft-tissues more readily. However, it is possible that matching a plain CT with a contrast-enhanced CT can affect DIR algorithm performance as the voxel intensities contained within the source and target VOIs may differ slightly due to the presence of contrast in one of the images. In addition, some of the contouring on the original pCTs was carried out with the aid of other image modalities, like PET and MRI (Table 3.2), which was not the case for the CTs that followed. All of these factors provide the RO with extra information to delineate the VOIs, which cannot be used by the DIR algorithm when deforming contours.

Table 5.1: Structure volume changes between CT scans for 5 patients (P1 to P5). P2 had 3 CT scans in total, resulting in 2 matches so that the pCT is matched to CT2 and CT2 is in turn matched to CT3. Negative values indicate a reduction in volume in the second CT scan.

Structure	**Patient**						**Mean**	**s.d.**
	P1	P2		P3	P4	P5	**(%)**	**(%)**
		CT2	CT3					
$t_{CT1\text{-}CT2}$ ***(weeks)***	6	4	2	5	4	4	4	1
$t_{F1\text{-}CT2}$ ***(weeks)***	3	1	3	3	3	1	2	1
ΔGTV (%)	-9.22	0.67	-1.83	-9.16	-4.45	-	-4.80	3.93
ΔCTV (%)	-7.52	0.13	-2.32	-4.13	-4.10	4.56	-2.23	3.80
ΔPTV (%)	-3.25	-1.72	1.98	-2.46	-4.08	2.75	-1.13	2.58
ΔBrain Stem (cm^3)	0.6	-1.5	-0.8	-0.2	0.9	-0.3	-0.2	0.8
ΔSpinal Cord (cm^3)	0.1	-0.3	0.0	-0.3	-0.5	-1.8	-0.5	0.6
ΔR Parotid (cm^3)	-1.1	-1.0	0.6	-1.7	-1.0	-1.4	-0.9	0.7
ΔL Parotid (cm^3)	-1.2	-1.5	-0.5	-1.0	-1.2	1.1	-0.7	0.9
ΔC3 (cm^3)	-0.3	-0.1	-0.5	-0.3	-0.2	0.1	-0.2	0.2

The mean and standard deviation of the percentage change in structure volumes for the patient cohort gives an indication of the degree of variation for each structure. However, the size of the cohort (5 patients, 6 matches) is small which makes the application of the mean and standard deviation to a general trend describing anatomical variation for head and neck cancer patients difficult. The means given in Table 5.1 serve to show significant changes for anatomy that is expected to change like the GTV, and smaller changes for anatomy not expected to change like the brain stem and C3 vertebra.

5.2. DIR Results of CT-CT Matches

For each of the six matches, the volumes delineated by the RO V_{RO} were compared to the volumes calculated by the DIR algorithm V_{DR} (Tables 5I – 5VI in Appendix 5A). The expression $V_{RO} \cap V_{DR}$ corresponds to the volume of overlap between the structures. This expression and the sum of volumes V_{RO} and V_{DR} are used to calculate the DSI as defined in Chapter 2, Equation 2.2.

Table 5.2: Summary of DSI scores for each structure in all clinical DIRs. Note that P2 had two sets of DIR, while P5 did not have a GTV delineated, but instead an added CTV60 and an added PTV60. *Not true means and standard deviations as they are only available for P5.

Structure	**Patient DSI**						**Mean**	**s.d.**
	P1	P2		P3	P4	P5		
		CT1-CT2	CT2-CT3					
GTV	0.82	0.79	0.85	0.87	0.97	-	0.86	0.06
CTV54	0.89	0.90	0.91	0.89	0.98	0.99	0.93	0.04
CTV60	-	-	-	-	-	0.99	0.99*	0.00*
CTV66	0.89	0.85	0.87	0.88	0.98	0.97	0.91	0.05
PTV54	0.93	0.92	0.94	0.91	0.99	0.99	0.95	0.03
PTV60	-	-	-	-	-	0.99	0.99*	0.00*
PTV66	0.92	0.87	0.88	0.90	0.99	0.99	0.93	0.05
Brain Stem	0.89	0.88	0.88	0.92	0.98	0.99	0.93	0.05
Spinal Cord	0.81	0.82	0.83	0.80	0.88	0.88	0.84	0.03
R Parotid	0.92	0.92	0.96	0.99	0.98	0.99	0.96	0.03
L Parotid	0.93	0.89	0.97	0.99	0.98	0.97	0.96	0.03
C3	0.76	0.76	0.79	0.81	0.75	0.93	0.78	0.07

The DSI scores were comparable across all matches with a range of 0.75-0.99 and average of 0.89 ± 0.05, suggesting high DIR accuracy.

Patient P2 was the only patient to have 3 CT scans in total,where the original planning CT was contrast-enhanced while CT2 and CT3 were not. This meant that the CT1-CT2 match for this patient was between a contrast-enhanced CT and a plain CT, while the CT2-CT3 match registered two non-contrast images. To determine whether there is a statistically significant difference in DIR performance between matching a plain pair of images and a contrast to plain registration, the DSI scores from both matches of P2 were compared (Table 5.2). The result of a paired, two-tailed Student's t-test showed that there is a statistically significant improvement in DSI scores for DIRs of non-contrast enhanced images versus a match between plain and contrast-enhanced images ($p = 0.007$). However, this result comes from the data of only 1 patient, so more patient data would have increased the reliability of this observation.

For patients P1-P4, the DSI scores for the C3 contours were unexpectedly lower than the overall average DSI score (Table 5.2). The discrepancies tended to manifest at the superior/inferior ends of the volume of interest (VOI). This suggests that the most likely reason for the discrpeancy is the effect of the interpolation between CT slices when reconstructing the 3D patient volume. That is, grey values are interpolated between each slice so that the same grey values are rendered in two adjacent slices, causing the algorithm to include/exclude them from the contour. In addition, the relatively small size of the VOI may have made it more prone to error because small discrepancies of a few voxels can have a bigger impact on smaller volumes than larger ones like the PTV. Figure 5.1 compares the DIR performance in the cases of patients P5 and P2, where the latter shows a discrepancy in the matching of the superior end of the C3 contours. P5's image shows greater separation between the vertebrae (bigger grey gaps) than P2's image, which may explain its higher DSI

score, as the contrast in grey values makes it easier for the algorithm to deform this part of the image.

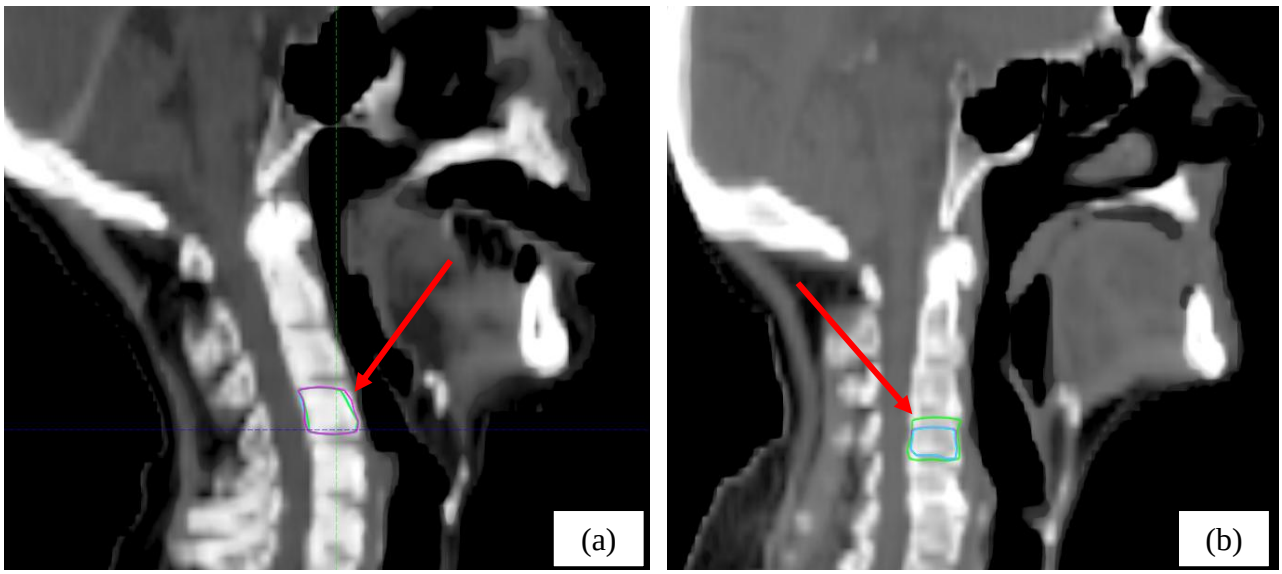

Figure 5.1: a) Sagittal slice of patient P5 after DIR where the source and target C3 contours are accurately matched, in magenta and green respectively. The cyan contour is the intersecting volume created by the Boolean operations tool. b) Patient P2 sagittal slice, showing mismatch between DIR-generated contour and Author-drawn contour of C3 in the DIR between CT2 and CT3. The cyan intersecting volume covers the deformed contour in magenta (cannot be seen).

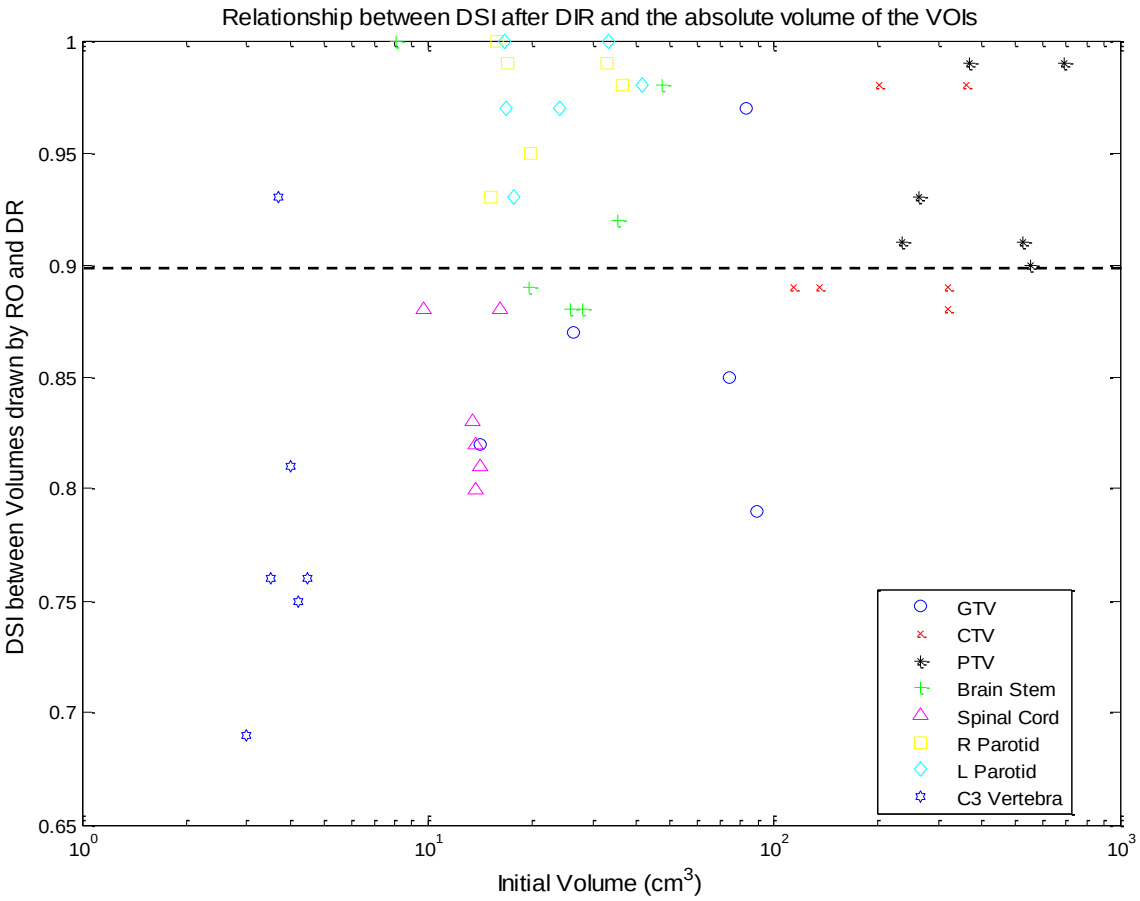

Figure 5.2: Plot showing the relationship between the volume of a structure and the observed DSI. The dashed line is the global average DSI for all VOIs.

Figure 5.2 illustrates the relationship between the initial volume of a structure and the observed DSI after DIR. The purpose of this graph was to determine whether small VOIs, such as the C3 vertebra, resulted in lower DSI scores than large structures like the CTV or PTV. It is possible that the smaller VOIs are more affected by discrepancies the size of a few voxels than larger structures, because the difference is divided by a smaller volume than for those of larger structures.

The graph shows marked variation in the volumes and DSI scores of the different VOIs, but there does not appear to be a correlation between the initial volume of the structures and the consequent DSI score from matching them, although an increase in the size of the sample would improve its reliability. Looking at the dashed line, small and large structures are

similarly represented above and below the average DSI. However, the graph does show different DIR performance for different VOIs not related to absolute volume. For example, the spinal cord DSI scores are all lower than the average, while the parotids and the PTV have higher scores.

The main factor contributing to the lower DSI scores for the spinal cord was an error in the Boolean operations tool of the TPS that drew the contour of the overlapping volume to be smaller than the actual volume of overlap between the structures (Figure 5.3). This underestimation therefore meant the numerator in the DSI equation had a smaller value (Equation 2.2) and hence a lower DSI.

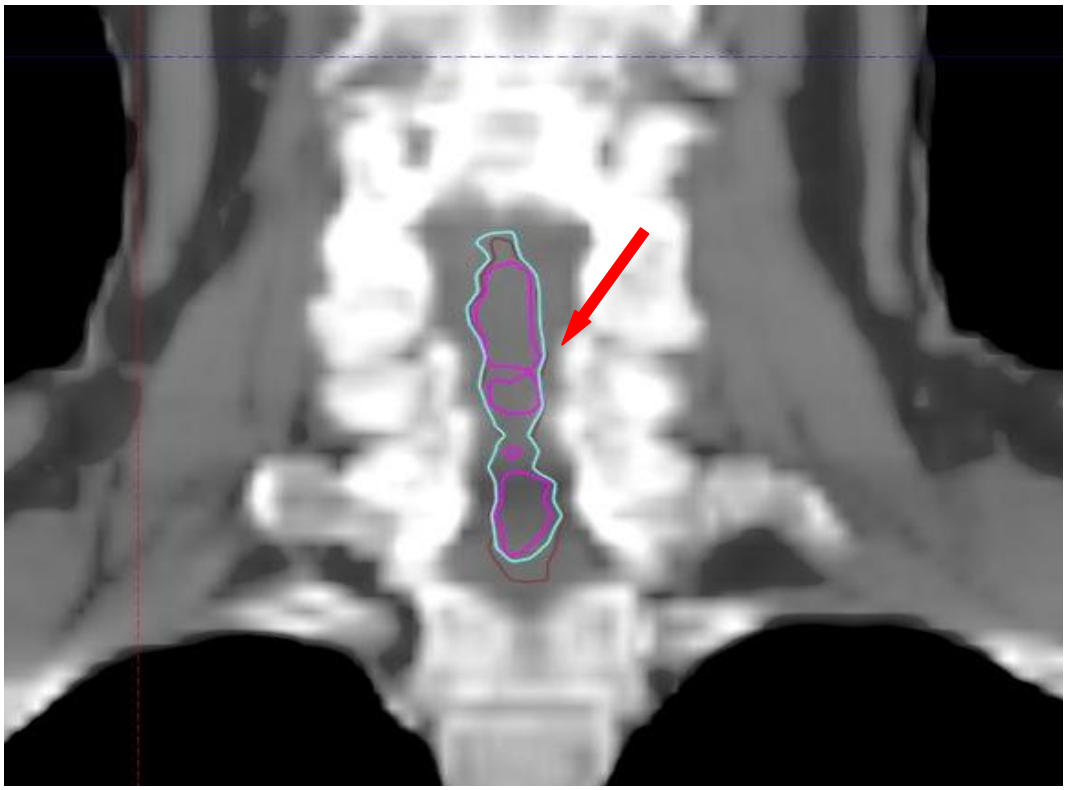

Figure 5.3: Coronal slice through the target image after DIR with the deformed spinal cord contour (brown) propagated from the deformed source image coinciding with the target contour (cyan). The overlapping volume (magenta) created by TPS is patchy and does not represent the true overlap.

5.3. DIR with Local ROI on CT-CT Matches

In this investigation the DIRs from Section 5.2, where a global ROI was applied, were repeated using ROIs limited to the PTV-region as defined in 3.1.1. The GTV, CTV, PTV, R Parotid, L Parotid and C3 vertebra were included in this local ROI and their DSI scores could be calculated and compared to the average DSI scores for the global ROI (Table 5.3). The PTV was chosen as the focus of the local ROI (Figure 5.4) for its central role in radiotherapy planning, as outlined in international guidelines [3].

Table 5.3: Comparison of average Global ROI and average Local ROI DSI scores.

	Global ROI DSI		**Local ROI DSI**		**ΔDSI**
Structure	**Mean (cm^3)**	**s.d. (cm^3)**	**Mean (cm^3)**	**s.d. (cm^3)**	
GTV	0.86	0.06	0.88	0.06	0.02
CTV	0.92	0.05	0.93	0.05	0.01
PTV	0.94	0.04	0.95	0.06	0.01
R Parotid	0.93	0.09	0.91	0.04	-0.02
L Parotid	0.93	0.08	0.89	0.04	-0.04
C3	0.78	0.07	0.74	0.08	-0.04

There was little difference in the overall mean DSI scores for the 2 ROI definitions, although it was observed that the DSI for the target volumes were slightly higher compared to the global ROI while the DSI for the other structures was slightly lower (Table 5.3). While this may indicate an improved DIR performance for the central regions of the image over the

periphery, the difference in DSI scores did not appear statistically significant, as confirmed by a Student's t-test ($p = 0.40$).

However, using a local ROI can have some benefits in situations where a physical change in a particular VOI is suspected. Carrying out DIR with a local ROI limited to that VOI and its immediate surrounding anatomy can give results optimised for this region without being compromised by other regions of the image. For example, patient P2's DSI score for the GTV in the CT1-CT2 match, was better than the DSI score achieved under the Local ROI boundaries ($DSI_{Local} = 0.89$ while $DSI_{Global} = 0.79$). This coincided with a note in the treatment record, where an RO had indicated that the "tumour has visibly grown" in the period between the two scans. Therefore, optimising a DIR to a given VOI using a local ROI can be beneficial in quantifying the change in this VOI and modifying the treatment plan accordingly.

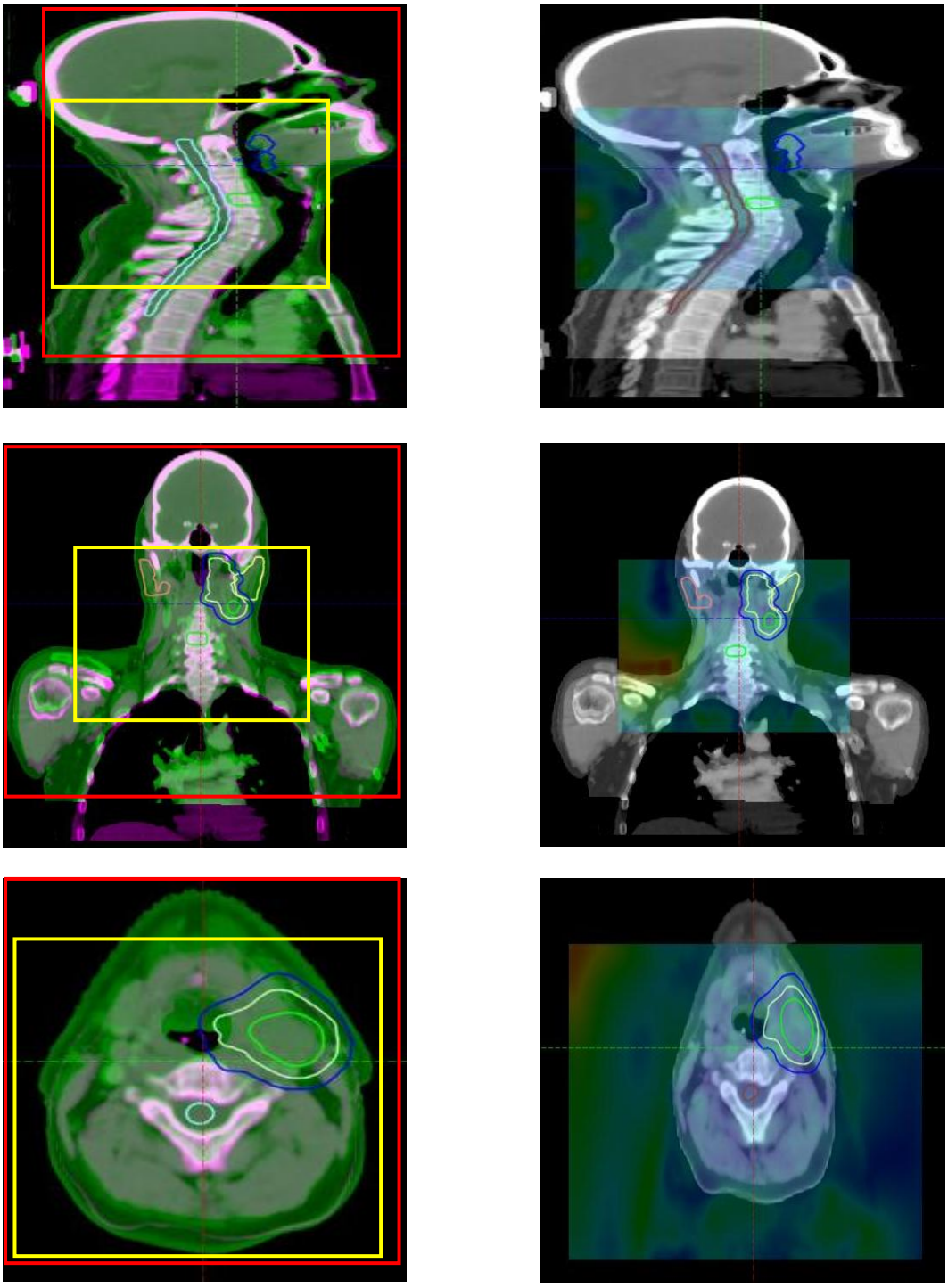

Figure 5.4: The relative sizes of the global (red) and local (yellow) ROIs are shown on the sagittal, coronal and axial slices of P1's registered source and target images on the left column. The results of DIR with the local ROI are shown on the right column. The colour intensity map representing the DVF produced inside the local ROI is superimposed on the images.

5.4. Qualitative Observations of DIR Performance

In general, the qualitative results showed good agreement between the reference VOI contours and the DIR-generated contours, while there were no major physically unrealistic deformations in the images (Figure 5.5 for example). The ROs approved all DIR-generated contours after slice-by-slice evaluation and amendment where appropriate. Most OARs required no or only minor modifications for clinical use. Most amendments were carried out for the GTVs (patients P1-P4) or the CTVs (P5 had no post-operative GTV). Superimposed DVF visualisations were used to identify apparent weight-loss and swelling that were corroborated by checking RO/nutritionist comments in the clinical notes. Not all qualitative results could be presented here so in the following two sections, a few examples are given to represent these qualitative observations.

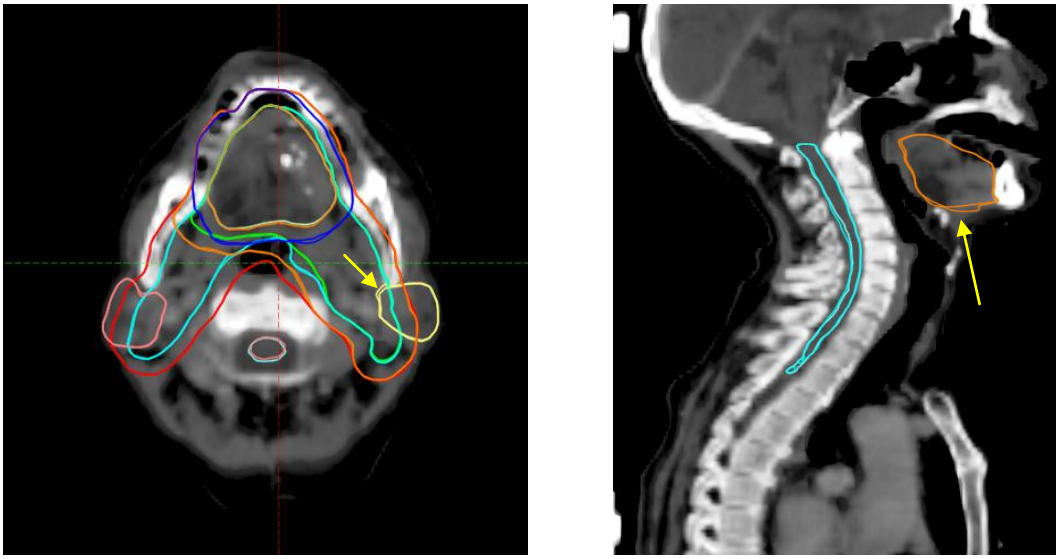

Figure 5.5: Patient P5 transverse and sagittal slices showing accurate DIR performance assessed qualitatively. Yellow arrows show minor adjustments made by the RO that resulted in a difference of the deformed and reference contour.

5.4.1. Contour Discrepancies

In most images the contours drawn by the RO and those resulting from DIR overlap showed good agreement between the DIR results and the reference contours. However, on some image slices, discrepancies can be observed (Figures 5.6-5.7) which highlight the limitations of the DIR generated structure.

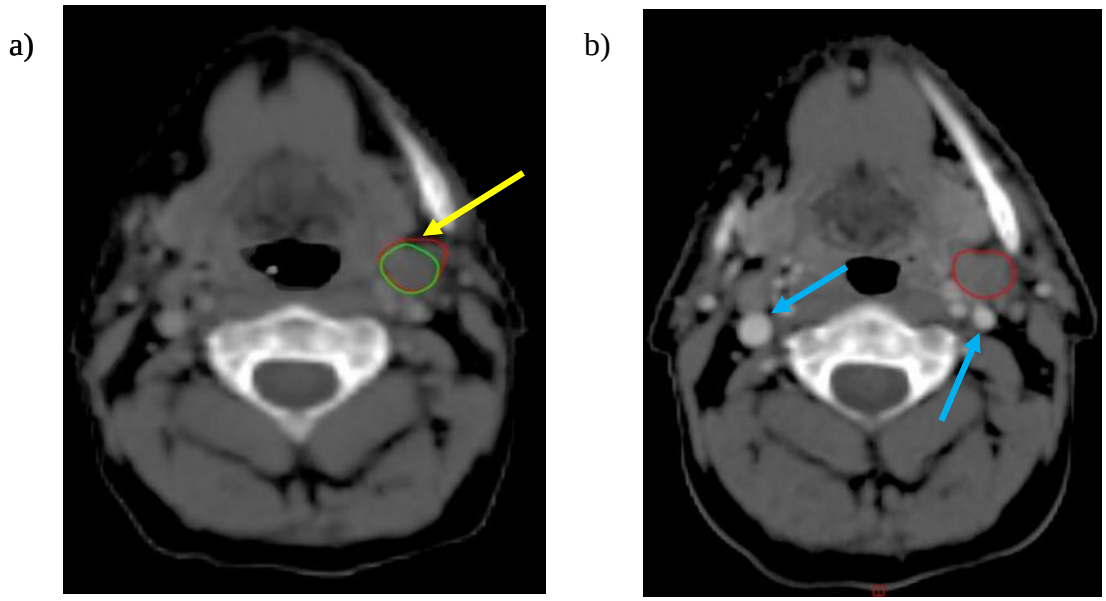

Figure 5.6: a) Transverse slice of P1's second CT scan showing the RO-drawn GTV (green) and the DIR-generated GTV (red) contours superimposed after DIR. contours. b) Same view of the patient anatomy in the original pCT image showing the GTV contour as it was drawn originally by the RO before DIR. Light blue arrows show some of the blood vessels enhanced by IV contrast.

Figure 5.6a displays an example of a mismatch between the RO's (green contour) and the algorithm's (red contour) delineation. It is apparent that for this GTV the algorithm attempts to follow the anatomy (light gray intensity) to its edge abutting the small cavity (dark region) that separates it from the jaw bone (bright white), while the RO chooses to limit the contour to a smaller volume (yellow arrow). Discussing this discrepancy with the RO revealed that the RO's decision was based on distinguishing anatomical details enhanced by the IV contrast in the original CT (Figure 5.6b). On the original pCT, the RO is able to see a lighter-intensity edge separating the GTV from neighbouring healthy soft-tissue that has very similar

grey values. However, no contrast was used in the second CT so that the GTV appeared to have the same voxel intensity as the neighbouring non-malignant tissue. Therefore, the algorithm, which can only use voxel intensities in its computations, incorrectly includes the extra tissue in the GTV contour.

It was also observed that a significant proportion of the structure volume discrepancies tended to manifest at the extremities of the VOIs (Figure 5.7). This is not to be confused with the situation where structures that are localised at the edges of the image (e.g. brain stem) are affected by a difference in the scan length and position of the registered images. One possible reason behind these discrepancies is that the image resolution in the axial direction is lower (3 mm) than the in-plane resolution (~ 1 mm), so voxels can be excluded/included depending on which slice they are happen to be in.

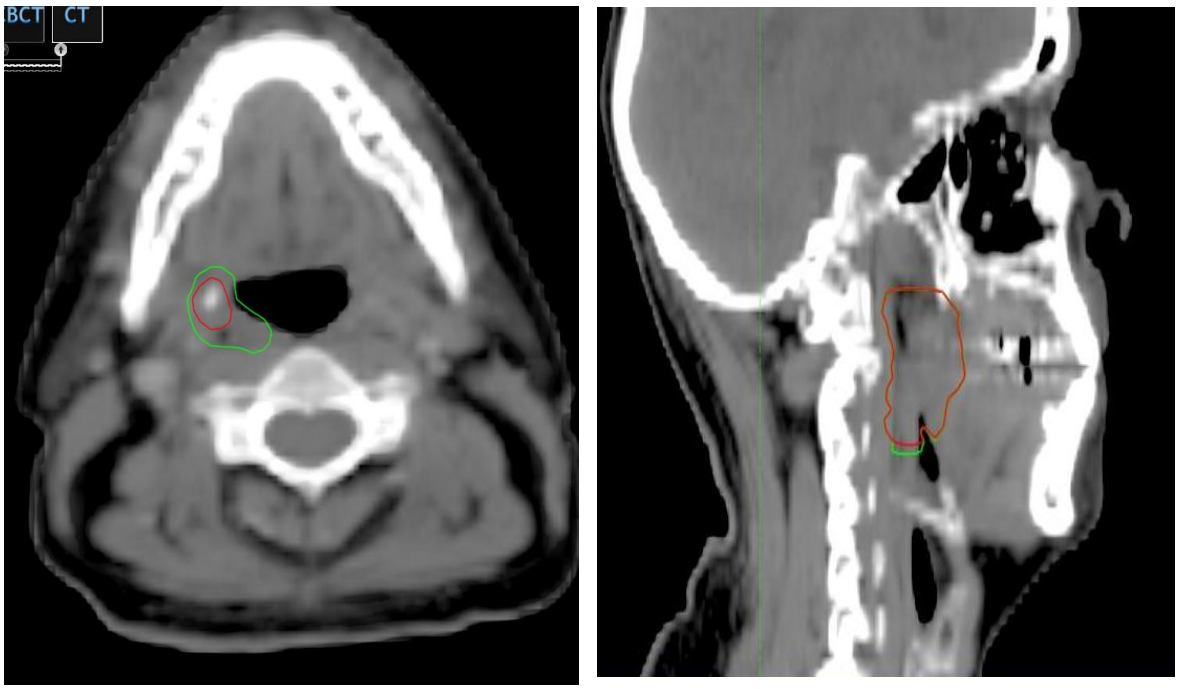

Figure 5.7: Transverse slice (left) and sagittal slice (right) of P3's second CT scan showing the RO's drawn CTV66 contour superimposed with the DIR generated contour, the discrepancy of the match is at the inferior side of the structure.

5.4.2. Deformation Vector Field Visualisation

Displaying the magnitude and/or direction of the deformation vectors aided in the identification of the sources of changes in some instances. The colour intensity map representing the DVF can be configured to display the magnitude of the deformation along a given cardinal direction. For example, it was noted in the patient record that P4 had experienced 7.4 kg of weight-loss during treatment between the 2 CT scans (Table 3.2), which can be correlated to the observed changes in the x-component of the DVF (Figure 5.8).

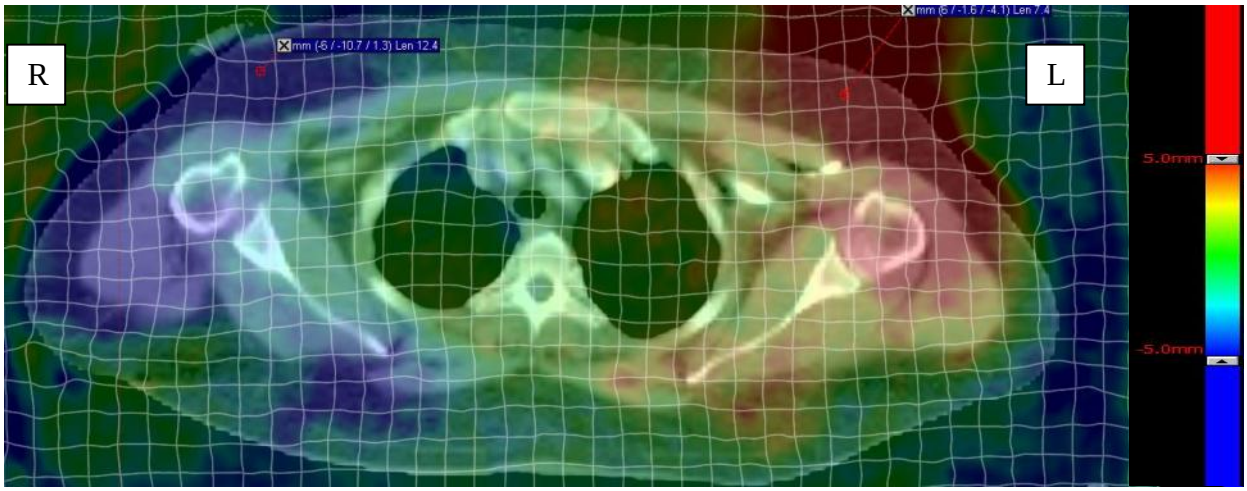

Figure 5.8: Transverse slice of P4's 2 CT scans superimposed after DIR overlaid with the sagittal component of the deformation colour intensity map. The patient lies in the head-first supine position so L and R indicate the left and right sides of the patient respectively.

In this example, the colour bar limits were set to ± 5 mm as this distance is representative of the minimum margin with which the PTV is expanded to cover the CTV as per the departmental protocol [49]. The co-ordinate system for the colour map in the sagittal direction displays voxel intensity movements towards the left of the patient as negative (blue), while voxel intensity movements towards the right are positive (red). The image is also overlaid with the deformation mesh that is meant to represent the deformation vector field (DVF) computed by the DIR algorithm. The patient's left is overlaid with red intensity that indicates a shift to the right (towards the spine), while the patient's right is overlaid with

blue that indicates a shift to the left. Therefore, this pattern of deformation can be liked to the weight-loss experienced by the patient in the 4-week period between the two scans.

In contrast, in Figure 5.9, P3's left side of the neck displays a blue (leftward) shift away from the spine while the right side of the neck displays a red (rightward) shift also away from the spine. Noting that the localisation of these shifts was limited to the neck only, indicates swelling or inflammation as the likely change rather than an overall weight change in the patient. Examination of the clinical record revealed a note by a clinician indicating the patient had mucositis (inflammation) in the neck region the day before the second CT scan.

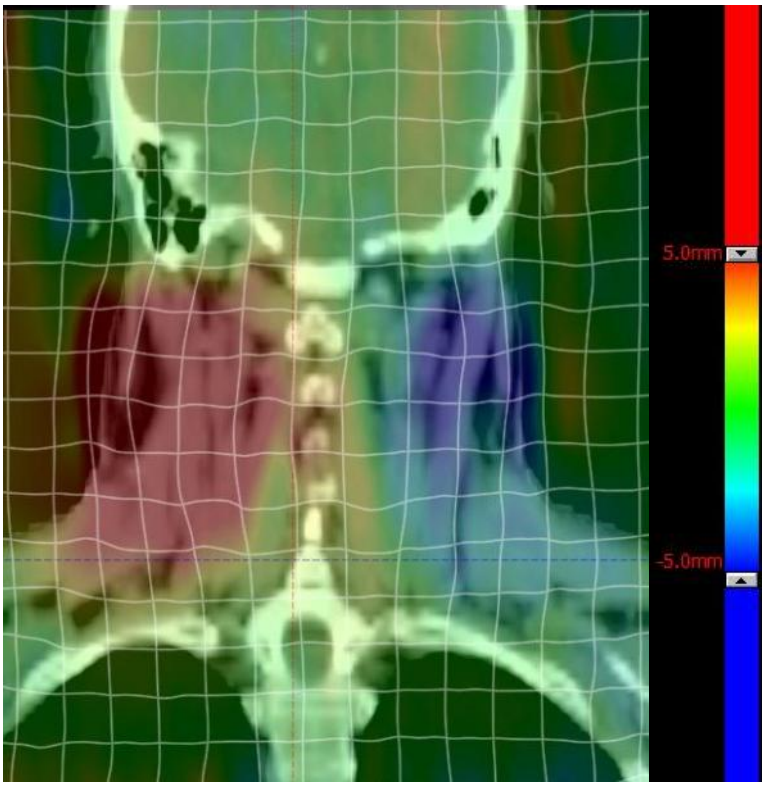

Figure 5.9: Coronal slice of P3's 2 CT scans superimposed after DIR overlaid with the x-component (sagittal) of the deformation colour intensity map.

On the whole, the quantitative and qualitative analysis agreed with information provided by the RO and indicated in the clinical record of the relevant changes.

Appendix 5A

Table 5I: Structure volume statistics and DSI scores comparing automatically deformed contours and Radiation Oncologist-drawn contours for Patient 1.

Patient 1 (P1)						
Structure	**V_{RO} (cm^3)**	**V_{DR} (cm^3)**	**V_{RO-DR} (cm^3)**	**$V_{RO} \cap V_{DR}$ (cm^3)**	**$V_{RO} + V_{DR}$ (cm^3)**	**DSI**
GTV	12.8	12.3	0.5	10.3	25.1	0.82
CTV54	190.2	189.2	1.0	168.0	379.4	0.89
CTV66	62.3	63.0	-0.7	56.0	125.3	0.89
PTV54	373.3	367.4	5.9	343.7	740.7	0.93
PTV66	131.3	133.4	-2.1	121.4	264.7	0.92
Brain Stem	20.3	18.3	2.0	14.9	38.6	0.77
Spinal Cord	14.1	13.6	0.5	8.5	27.7	0.67
R Parotid	18.8	18.0	-0.8	16.9	36.8	0.92
L Parotid	18.7	19.2	-0.5	18.0	38.9	0.93
C3	3.2	4.2	-1.0	2.8	7.4	0.76

Table 5II: Patient 2 structure volume statistics and DSI scores for DIR of CT1 to CT2.

Patient 2 (P2) CT1-CT2						
Structure	**V_{RO} (cm^3)**	**V_{DR} (cm^3)**	**V_{RO-DR} (cm^3)**	**$V_{RO} \cap V_{DR}$ (cm^3)**	**$V_{RO} + V_{DR}$ (cm^3)**	**DSI**
GTV	105.7	89.6	16.1	77.0	195.3	0.79
CTV54	443.5	421.4	22.1	389.2	864.9	0.90
CTV66	221.5	191.7	29.8	174.9	413.2	0.85
PTV54	754.5	731.0	23.5	680.1	1485.5	0.92
PTV66	362.7	318.0	44.7	297.0	680.7	0.87
Brain Stem	25.8	25.7	0.1	22.7	51.5	0.88
Spinal Cord	13.5	12.6	0.9	9.1	26.1	0.70
R Parotid	15.2	15.4	0.2	14.1	30.6	0.92
L Parotid	17.7	17.1	-0.6	15.5	34.8	0.89
C3	4.1	3.0	1.1	2.7	7.1	0.76

Table 5III: Patient 2 structure volume statistics and DSI scores for DIR of CT2 to CT3.

Patient 2 (P2) CT2-CT3						
Structure	**V_{RO} (cm^3)**	**V_{DR} (cm^3)**	**V_{RO-DR} (cm^3)**	**$V_{RO} \cap V_{DR}$ (cm^3)**	**$V_{RO} + V_{DR}$ (cm^3)**	**DSI**
GTV	74.6	74.2	0.4	63.2	148.8	0.85
CTV54	433.2	433.1	0.1	396.3	866.3	0.91
CTV66	175.3	190.1	-14.8	158.7	365.4	0.87
PTV54	746.3	745.5	0.8	698.9	1491.8	0.94
PTV66	331.8	312.0	19.8	282.4	643.8	0.88
Brain Stem	25.0	24.5	0.5	21.7	49.5	0.88
Spinal Cord	13.5	13.4	0.1	9.5	26.9	0.71
R Parotid	15.8	15.3	-	14.9	31.1	0.96
L Parotid	16.6	16.0	-0.6	15.8	32.6	0.97
C3	4.9	2.9	2.0	2.7	7.8	0.69

Table 5IV: Patient 3 structure volume statistics and DSI scores.

Patient 3 (P3)						
Structure	**V_{RO} (cm^3)**	**V_{DR} (cm^3)**	**V_{RO-DR} (cm^3)**	**$V_{RO} \cap V_{DR}$ (cm^3)**	**$V_{RO} + V_{DR}$ (cm^3)**	**DSI**
GTV	23.8	22.9	0.9	20.2	46.7	0.87
CTV54	163.5	163.0	0.5	144.7	326.5	0.89
CTV66	58.5	62.3	-3.8	53.3	120.8	0.88
PTV54	346.4	347.7	-1.3	316.1	694.1	0.91
PTV66	116.2	124.8	-8.6	108.7	241.0	0.90
Brain Stem	20.7	21.2	0.5	17.6	56.7	0.84
Spinal Cord	13.1	13.8	0.7	8.7	41.3	0.65
R Parotid	27.7	27.4	0.3	27.3	55.1	0.99
L Parotid	28.7	29.6	0.9	28.8	58.3	0.99
C3	3.8	3.4	0.4	2.9	7.2	0.81

Table 5V: Patient 4 structure volume statistics and DSI scores.

Patient 4 (P4)						
Structure	**V_{RO} (cm^3)**	**V_{DR} (cm^3)**	**V_{RO-DR} (cm^3)**	**$V_{RO} \cap V_{DR}$ (cm^3)**	**$V_{RO} + V_{DR}$ (cm^3)**	**DSI**
GTV	79.3	77.5	1.8	76.4	156.8	0.97
CTV54	540.5	533.1	7.4	524.8	1073.6	0.98
CTV66	148.6	146.9	1.7	145.2	295.5	0.98
PTV54	1050.2	1050.7	-0.5	1040.9	2100.9	0.99
PTV66	276.4	276.9	-0.5	275.0	553.3	0.99
Brain Stem	48.3	48.5	-0.2	47.6	96.8	0.98
Spinal Cord	14.0	14.6	-0.6	12.6	28.6	0.88
R Parotid	32.3	32.0	0.3	31.6	64.3	0.98
L Parotid	38.6	38.4	0.2	37.9	77.0	0.98
C3	4.8	3.7	1.1	3.2	8.5	0.75

Table 5VI: Patient 5 structure volume statistics and DSI scores.

Patient 5 (P5)						
Structure	**V_{RO} (cm^3)**	**V_{DR} (cm^3)**	**V_{RO-DR} (cm^3)**	**$V_{RO} \cap V_{DR}$ (cm^3)**	**$V_{RO} + V_{DR}$ (cm^3)**	**DSI**
CTV54	306.5	308.3	1.8	303.2	614.8	0.99
CTV60	225.1	226.3	1.2	222.9	451.4	0.99
CTV66	62.1	61.8	-0.3	60.4	123.9	0.97
PTV54	591.4	592.3	0.9	587.2	1183.7	0.99
PTV60	411.5	412.7	1.2	409.2	824.2	0.99
PTV66	124.5	125.5	1.0	124.0	250.0	0.99
Brain Stem	7.8	7.8	0.0	7.8	15.6	1.00
Spinal Cord	7.9	8.5	0.6	7.2	16.4	0.88
R Parotid	14.9	14.7	-0.2	14.6	29.6	0.99
L Parotid	17.7	17.5	-0.2	17.1	35.2	0.97
C3	3.4	3.5	0.1	3.2	6.9	0.93

6. Discussion & Conclusions

6.1. Initial Investigations

6.1.1. DIRART Algorithm Selection

After the initial investigations, the Horn-Schunck Optical Flow algorithm (HS) was selected for the subsequent investigations out of the 4 algorithms from DIRART as described in Section 3.3.1. HS did have the *r* score closest to 1 out of the four algorithms tested, although the combined Lucas-Kanad and HS (LK+HS) and Demons with Elastic Regularisation (D+ER) algorithms had similarly high scores (Table 4.1). Nonetheless, the HS algorithm also gave the closest match between the deformed source image and the target image (Figures 4.1-4.4) qualitatively.

The visualisation of the deformation vector field (DVF) highlighted a number of inaccuracies in the algorithms' interpretation of the volume change exhibited by the images. Firstly, all algorithms displayed non-zero vectors in regions of the images where no change was induced (regions circled in red in Figures 4.1-4.4). A possible explanation for these errors may be a movement during the replacement of the breast attachment, which was most likely a rotation about the coronal anterior-posterior axis as evidenced by the same distribution of vectors seen in the axial views of all 4 DIRs. A rigid registration process allowing rotation as well as translation would compensate for this shift; however, DIRART rigid registration only allows translations [34]. Consequently, the DIR algorithms would attempt to compensate for this shift by compromising the deformation to give rise to these vectors.

The second DVF inaccuracy was seen in the curving vectors to the top-right of the breast, and proximal to the edges of the registration region of interest (ROI) (Figures 4.1-4.4). To rule

out that the incorrect vectors were the result of the left ROI boundary's proximity to the VOI, the deformations were repeated with a wider ROI but the resultant DVF remained the same. Furthermore, it recurred across all four algorithms and hence it was not algorithm specific. A possible explanation for this error could be the shape of the removed breast section that has a markedly more elongated mould in the relevant region of the VOI when compared to the smaller breast section still attached to the phantom. Therefore, the removal of a substantial number of voxel intensities in a sharply tapering region of the VOI may have presented the DIR algorithms with a large discontinuity that could not be accurately processed.

Furthermore, the HS algorithm also had a relatively fast computation speed that contributed to its selection (Table 4.1). Although registration speed was not a major concern in this study, it can be an important issue when designing a clinical strategy for online adaptive radiotherapy because this will add to the overall treatment time [5, 50].

In a study by Yeo *et al* on DIR and deformable phantoms, a number of DIRART's algorithms, including the HS and Fast Demons algorithm were compared and analysed for their performance. The results of that study also indicated that the original HS algorithm was the most reliable [22].

Kirby *et al.* evaluated the accuracy of 11 different DIR algorithms using a deformable phantom that represents the axial plane of pelvic anatomy [23]. Nine of the algorithms tested came from DIRART including the HS, FF, and D+ER algorithms investigated in this study. Quantitatively, they assessed the DIR algorithms using 3 error-measuring parameters and 2 similarity measures, one of which was the Dice Similarity Index (DSI). Comparing the 3 algorithms common to both studies, they found that the D+ER algorithm gave the highest DSI score (0.93) followed by the HS algorithm (0.74), while the FF algorithm gave the poorest results (0.68). While the observed performance of the 3 algorithms was similar to those in this investigation, a direct comparison is difficult for several reasons, namely: (I)

Kirby *et al.* used different assessment criteria in their study (mean absolute distance and displacement error); (II) Kirby *et al.* used an internal VOI representative of the rectum, a different anatomical site; (III) the deformation simulated by Kirby *et al.* was only in 2 dimensions. Furthermore, the study concluded that different DIR algorithms yield very different results for a given case and that a DIR algorithm that is optimised for one application could yield large errors for another [23].

6.1.2. CT Slice Resolution

This part of the study was designed to assess the impact of CT slice thickness on contour volume delineation in the planning system, which in turn can affect the ability of the DIR algorithm to determine the contour boundaries. CT slice thickness is known to affect an image's low-contrast resolution as well as its susceptibility to image artefacts due to partial volume effects, and blurring due to noise [45].

There was no straightforward explanation for the observed relationship between the CT slice resolution of the scan and the volume of interest discrepancy (Figure 4.5). Multiple sources of uncertainties inherent in the CT slice uncertainty contribute to DIR algorithm performance.

Firstly, slice-ambiguity can arise from the arbitrary selection of the initial start position of the CT scan; that is, the edges of the actual VOI may fall at the beginning or middle of the first slice. Due to the inherent averaging of tissue density between slices, the most distant parts of the VOI may either be just included or excluded in a specific CT slice depending on the start position of the CT scan, affecting the observed magnitude of the VOI.

Secondly, there is an uncertainty contribution from automatic contouring due to its dependence on the Hounsfield Unit (HU) ranger to delineate the bone-equivalent discs comprising the VOI. The HU ranger is set to use a wide span (200-1700 HU) for contouring bony anatomy because there is a large range of bone density clinically, while the CIRS

phantom's bone-equivalent tissue is one homogenous density [2, 51]. Therefore, automatic contouring can include some of the neighbouring soft-tissue equivalent material of the phantom body, especially at the interface where some blurring occurred. Consequently, the automatically contoured VOIs overestimated the actual volume of the VOI in both the source and target images for all CT slice resolutions (Section 4.1.2 and Table 4.2).

There are a number of limitations in this investigation that can be minimised in a future repetition of this study. Firstly, the number of scans at each resolution needs to be substantially increased in order to arrive at a more statistically definitive determination of the relationship between CT resolution and volume error. Secondly, the choice of the starting bed position from which the phantom was offset for subsequent scans was arbitrary and should have been chosen based on the location of the VOI within the phantom so that the bed position offsets can be correlated to the region of the VOI with which they coincide. Thirdly, the minimum required HU range for automatic contouring should be incorporated in this investigation to decouple its contribution to contouring uncertainty from the uncertainty due to slice resolution.

A few studies had investigated the effect of CT slice resolution on contouring in radiotherapy planning. While some showed that differences arising from slice thickness for some organs were statistically insignificant, other studies showed consistent volume underestimation for a particular organ on a given slice thickness (5 mm) compared to a thinner slice (3 mm) [51, 52].

6.2. Similarity Measures

There were two similarity measures (SMs) explicitly used in this study to quantitatively assess the performance of the DIR algorithms, Pearson's product moment correlation coefficient (r) and the dice similarity index (DSI). However, due to the technical problems

encountered in running DIRART, the use of *r* was limited to the gross volume change investigation (Section 4.2).The SmartAdapt v10 (SA) DIR programme did not provide any similarity measure information so DSI was chosen for the simplicity of its application based on the drawing of volumes facilitated by the treatment planning system (TPS). Furthermore, DSI is a transparent SM with no hidden parameters and a finite interval (0 to 1) that is intuitive. Its main disadvantage is the lack of information on the agreement of voxel intensities in a given VOI [18].

6.3. Investigations Using Phantoms

The main benefit of the phantom investigations was the ability to control the induced changes and objectively assess the accuracy of the various DIR algorithms.

6.3.1. Volume Change Using the RANDO Phantom

This investigation simulated a change in shape and volume (Section 3.3.2). The quantitative evaluation produced high SM scores with an average DSI and *r* scores of 0.90 ± 0.02 and 0.85 ± 0.02 respectively. While not directly comparable, the DSI score was closer to 1 than the *r* score, as expected, due to the presence of the air gap in the VOI (Figure4.6), which affects voxel-intensity similarity but not contour volume. The uncertainty in the DSI score was within what was expected after taking into account the propagation of contouring uncertainty (Section 4.2), which meant that the Boolean operation used to automatically calculate the overlap between the VOIs did not appear to introduce discernible errors in DSI calculation.

Qualitatively, both the Modified Demons algorithm in SmartAdapt (SA) and the HS algorithm in DIRART did not result in large discrepancies between the deformed and target images that would lead to clinically relevant errors (Figure 4.6).

Both DIR algorithms deform the air gap between the two sections of the breast (Figure 4.6) in such a way that the size of the target volumes is conserved due to their smoothness constraint function (Section 2.2, Equation 2.1). This function is intended to prevent unrealistic deformations but actually results in a physically unrealistic VOI in this example because the algorithms are forced to retain the air gap.

The results of this phantom investigation can be compared to the clinical scenario of patient weight change. That is, even though the deformed and target VOIs were closely matched, the error in deforming the air gap *inside* the VOI did not affect the volume matching as a whole. Similarly, in a case of patient weight change it would be relatively easy to determine which regions of the patient's head and neck gained/lost mass, but this will not reveal the relative changes in the individual organs at risk (OARs) or target volumes (TVs).

6.3.2. Local Volume Change + Translation in the CIRS Phantom

This investigation simulated an increase in the size of an internal VOI using a rigid phantom (Section 3.3.3). The DIR results for this investigation were poorer quantitatively (DSI = 0.69 ± 0.02) and qualitatively (Figure 4.7) when compared to the DIR results of the previous investigation. The main difference between the volume change simulated in this investigation and the one simulated in the RANDO phantom (Section 6.4.1) is that here the VOI is an internal volume inside the phantom. Most clinical scenarios of DIR would involve dealing with deformations due to changes that occur in internal VOIs such as patient movement within the immobilisation mask, GTV shrinkage, or inflammation. This experiment also underscored the problem of missing/added tissues, i.e. the difference of the information in the source and target images that a DIR algorithm must deal with when computing the deformation field.

Looking at Figure 4.7, the source image was insufficiently deformed along the axial direction, i.e. the direction of the volume change and translation. This constituted the main source of error. There are a number of possible reasons for this error: (I) the presence of gaps between the inserts that are due to the uneven surface of the putty filling of the discs. Essentially, the presence of low intensity voxels between groups of high intensity voxels in the axial direction may confound the algorithm's optimisation between the similarity function and smoothness constraint (Equation 2.1). However, this hypothesis does not support the partial deformation of the superior insert in Figure 4.7d. (II) The number of levels of the multi-resolution pyramid used in the Modified Demons algorithm (Section 2.2.3) was too low. Since the deformation appeared to have been terminated before fully deforming the source image it is possible that the addition of another resolution level to the process would improve the DIR. However, SA does not allow the modification of the algorithm parameters and so it was not possible to test this hypothesis. (III) The CT resolution in the axial direction (3 mm slice thickness) is lower than the in-plane resolution (≈ 1 mm x 1 mm), which can arguably explain the better deformation of the VOI in the sagittal and coronal planes.

6.3.3. Rotation and Deformations in the Clay Phantom

The rod of clay used as the VOI in this investigation was attached to a rigid solid water base (Figures 3.9, 4.8-9, 4.11) and configured to simulate a rotation and deformations.

Rotation

This experiment tested the boundary conditions of the Modified Demons algorithm when dealing with a relatively significant change in the VOI and the results highlighted how ROI selection affects the outcome of the DIR (Figure 4.8).

DIR performed with an ROI localised on the VOI overcame the error introduced by using a global ROI encompassing the whole phantom (Figure 4.8). This was the result of the

competing functions in the DIR algorithm optimisation process: similarity and smoothness. When a global ROI is used the unchanged solid water base constitutes the largest grouping of homogeneous voxel intensities (approximately 0 HU), while the VOI (voxel intensities ≈ 1800 HU) makes up the other smaller grouping. Therefore, when the algorithm searches for the optimal solution to the deformation problem, the smoothness constraint restricts algorithm convergence to a sub-optimal similarity value because the change in the VOI is relatively large while the solid water base remains the same. A local ROI focuses the DIR on the VOI only and hence results in a more accurate deformation.

Deformations

This part of the investigation grouped two experiments together, 'bending' and 'pressing' (Section 4.4.2), which did not involve the simulation of a volume change and provided the best phantom DIR results both qualitatively (Figure 4.9) and quantitatively (Tables 4.7 and 4.8). However, the deformed contours in both experiments were of a smaller volume than the target image and source image (Tables 4.6 and 4.7), i.e. showing a false, albeit small volume decrease of 0.4 cm^3 and 0.1 cm^3 respectively.

Primarily, the high accuracy of these results can be attributed to the homogeneity of the VOI (no air gaps), i.e. a contiguous group of uniform voxel intensities in a VOI is less likely to compromise the algorithm's smoothness constraint. Additionally, the accuracy of the results can be partly attributed to the smaller displacement of the VOI and its unchanging volume, which means the algorithm can more accurately compute the deformation vectors that would transform the source image to match the target image [37]. That is, all voxel intensities present in the source image are present in the target image and so the algorithm is more likely to map a given voxel intensity to its counterpart because the same tissues are represented in both.

Irregular Volume Change (Cutting)

The cutting experiment involved a volume decrease of the rod by removing approximately 40% of its volume (Section 4.4.3). The final results were of a high accuracy (Figure 4.11, DSI = 0.91± 0.03) and resulted from using a local ROI, as initially a global ROI resulted in a physically aberrant result (Figure 4.10) for the same reasons discussed for the rotation experiment. This experiment gave the best DIR performance out of all volume change experiments, which can be most likely attributed to the higher homogeneity of the VOI, unlike the air gaps present in the RANDO and CIRS phantom VOIs.

Like the CIRS phantom investigations, contouring uncertainty was evident in the delineation of the VOIs in this experiment, but it did not significantly affect the quality of the deformation (Table 4.9). In all the phantom investigations the contouring uncertainty ranged from 0.5-2.9%, with no major difference between completely automatic contouring, as done for the Clay phantom (average of 1.8%) and the combined automatic and manual contouring used in the CIRS and RANDO investigations (average of 1.9%). However, the applicability of this range of contouring uncertainty to clinical contouring is not straightforward. The contouring of the TVs and OARs in the clinical cases was performed manually by the RO, which entails added uncertainty due to observer subjectivity, and intra-observer variability when it comes to drawing the same structures on new CT images.

6.4. Clinical Results

Ten clinically important structures were selected for assessments of the DIR accuracy in a clinical setting (Section 3.1.1), which can be divided into TVs and OARs.

Target volumes are the object of the prescribed radiation dose and include the GTV, clinical target volume (CTV), and planning target volume (PTV) [43]. These volumes have different consequences for DIR: (I) the GTV is often a physical volume and constitutes the whole

tumour visible in the image, which means that any changes it undergoes can be readily mapped using DIR. However, if the tumour was surgically removed, only a CTV is contoured. (II) The CTV on the other hand is theoretically a physical volume but practically it is a quasi-geometrical volume, i.e. it is a contour around the GTV that is meant to cover any microscopic cancerous tissue that cannot be seen on the image [4]. For the patients in this study the CTVs were contoured according to international guidelines on delineation of TVs, which use visible physical anatomy that are at risk of malignant invasion as CTV boundaries and/or use an expanded margin around the GTV [53]. (III) The PTV is a purely geometric structure that is an expansion of the CTV using a PTV margin. Its purpose is to ensure adequate coverage of the CTV in case of setup uncertainties and patient motion [4]. Changes in the GTV/CTV or surrounding physical volumes can indirectly affect the DIR-generated PTV; for example, GTV shrinkage can cause inward contraction of surrounding tissues that appears as PTV shrinkage after DIR.

In contrast to TVs, OARs are objects of avoidance and targets of sparing from radiation dose as much as reasonably achievable. They generally have smaller delineation uncertainties because they have clearly defined boundaries so their delineation is less subjective than TV delineation and should result in better agreement between DIR-generated and RO-drawn contours. As an addition, the third cervical vertebra (C3) was contoured to observe the accuracy of DIR on a structure not expected to deform.

6.4.1. Anatomical Changes

In Section 5.1, the volumes of each structure in the source and target images were compared to determine the extent of anatomical variation in the time between CT scans.

To begin, it is useful to look at the imaging timeline for the patients in this study (Tables 3.2 and 5.1). For patient P2 there were two CT scans in addition to the original planning CT taken over a period of 6 weeks; for the other 4 patients a second CT scan was taken 4-6

weeks after the original (Table 5.1). For all patients, treatment began 2-3 weeks from the day of the initial planning CT (Table 3.2). Looking at range of anatomical variations, two observations can be discussed: (I) With the exception of P2 all patients showed a decrease in the volume of the GTV/CTV, which can be attributed to positive patient response to the treatment. On average, the second CT for each of these patients was taken 3 weeks after the first treatment fraction (Table 5.1). This agrees with similar observations of GTV/CTV changes in head and neck radiotherapy studies that were considered candidates to benefit from DIR and adaptive radiotherapy (ART) to improve treatment outcome [5].

(II) While P2's VOIs did not present any significant changes in magnitude between CTs, there were systematic drifts in the patient set up during treatment. Comparison of routine cone-beam CT images of this patient to the original planning CT (pCT) were used to establish that this was due to changes in the patient's head and neck position within the immobilisation mask. In this case DIR was used to account for shifts in the positioning of VOIs and not morphological changes in anatomy. While regular cone-beam CT imaging is often sufficient to detect such changes for patient set-up, DIR can also be helpful in quantifying these changes. For example, deforming the contour of the PTV can be used to work out how much of the volume is no longer adequately covered by the treatment.

(III) For manual contouring, intra-observer variability can play a major role in the variation of the contour volumes, as was found in this study. Initially, the delineation of the parotid glands of P1-P3 varied considerably between CT scans, even though these OARs were delineated by the same RO in each case. However, in the course of reviewing the DIR results of these patients with the RO, it became apparent that these variations were due to inconsistent contouring, which were later corrected by the clinician for the final analysis. It is often difficult to identify and characterise all instances of intra-observer variability in a given image set [18], but a number of methods have been proposed to mitigate the effects of intra-

observer variability such as analysis of variance (ANOVA) [30] or joint delineation review sessions and application of multimodality imaging [54]. For example, Mencarelli *et al.* sought to determine the impact of human observer variation on the validation of DIR using the ANOVA method [30]. In that paper, the ANOVA method was applied to calculate the random variation of an individual observer, by examining the variations in the differences between the measurements of the observers. This approach was tailored to the point-based DIR (Section 2.1.1) method used in their study, and while this can be modified for voxel-intensity based DIR such as that used here, it was beyond the scope of this study.

6.4.2. DIR Performance in CT-CT Clinical Matches

In this part of the study, the accuracy of DIR performed using the Modified Demons algorithm in SA was evaluated by comparing VOIs checked and approved by an RO to the DIR results.

The VOIs generated by DIR matched the RO-drawn contours with sufficient accuracy for clinical use with minor or no corrections. DIR algorithm performance was on the whole equal for the different TVs and OARs. As illustrated in Figure 5.2, there was no apparent correlation between a structure's initial volume and how accurately the algorithm deforms it.

There were a few examples of unexpectedly low DSI scores. The overlapping volumes of the C3 vertebrae and spinal cords on the source and target images showed an error in the creation of the volume of intersection between the DIR-generated and RO-drawn contours (Figures 5.1, 5.3). The source of these errors was the Boolean operations tool that finds the intersection of two contours, which may be due to CT slice thickness uncertainty. Indeed, both the spinal cord and vertebrae are longitudinal structures that extend axially and hence were more susceptible to errors due to the lower resolution (3 mm) in the axial direction.

Some DIR studies have reported that the agreement of DIR-generated VOIs with physician drawn VOIs appeared to be dependent on the structure of interest. DIR-generated OARs were found to be more accurate than DIR-generated TVs due mainly to the subjective nature of target definition in radiotherapy planning [18, 55]. Conversely, the results of this study showed that DIR-generated target volumes, specifically CTVs and PTVs, gave relatively high DSI scores (Table 5.2). The reason for this apparent contradiction lies in the method of delineation of the volumes of interest in this study.

In this study, the VOIs modified by the RO after DIR are used as the reference with which to compare the deformed structures. The drawback for this approach is that the RO's contouring decisions can be influenced by the result of the deformation and hence skew the contouring decision in favour of agreement with the result of the DIR. Ideally, reference structures would be manually contoured on the target image by the RO in the same manner as the initial planning contouring, and then compared to the deformed structures. However, the contouring approach where the RO is blind to the deformation results is more time consuming and is still subject to intra-observer variability that would remain largely unknown without rigorous analysis. Therefore, for more subjective delineations, an RO is less likely to modify a DIR-generated contour readily propagated onto the target image, especially if there are no obvious clinically unrealistic delineations.

Another possible reason for the relatively high scores for the CTVs and PTVs is the large size of these VOIs. As mentioned earlier, mismatches might have a larger impact on small volumes closer to the voxel size than on large volumes.

6.4.3. DIR Using Local ROI

The phantom investigations results showed that the choice of the dimensions and location of the ROI boundaries for DIR can greatly affect the outcome of the deformation (Sections 6.4.2-3). Hence, the decision was made to investigate the effect of using DIR with a local ROI on clinical images and comparing the results to the default global ROI used originally. To avoid biasing the results, the same rigid registration used for the global ROI was taken as the starting point for the DIR using the local ROI.

A global ROI that encompasses the whole of the overlapping region of the registered images was chosen as the default approach to carrying out a registration (Section 3.1). The reason for this approach was that in a clinical situation not all the anatomical changes a patient had undergone between scans are known. Therefore, in order to account for all the changes, the images are matched as a whole. On occasion, the RO may want to specifically assess the change in a given structure of interest, such as the planning target volume (PTV), without deforming the whole image. In such cases, a local ROI centred on that structure and its immediate surrounding anatomy can be used, which might give a better quality match for the PTV than a DIR using a global ROI that optimises the deformation as a whole and thus compromises on the matching of the PTV.

As shown in Section 5.3, there was no statistically significant difference between using the global and local ROIs (Student's t-test, $p = 0.40$). Although, it appears that the average DSI scores of the TVs improved slightly when using the local ROI, while the average DSI scores of the C3 vertebra and the OARs appeared to deteriorate by a similar degree. A possible explanation is one related to the relative sizes of the VOI pairs being compared, where the smaller C3 and OARs are more prone to mismatches of a few voxels than the larger TVs. That is, although the DIR algorithm does not consider the contours when computing deformations (only voxel intensities), a small discrepancy in computing the deformation

vector field (DVF), e.g. due to CT slice thickness, can be propagated to the contour when the DVF is used to deform it.

6.4.4. Qualitative Assessment

Qualitative evaluation of the DIRs echoes the quantitative assessment and shows similarly accurate matching between the deformed and target VOIs. However, qualitative appraisal of the images also revealed additional information about DIR performance.

One of the main limitations of DIR is its sole reliance on voxel intensities and algorithms that are not able to relate anatomical information from pathology and histology as human observers. Consider the example in Figure 5.6; the DIR-generated GTV contour did not match the RO-drawn GTV because the algorithm attempted to contain part of the anatomy adjacent to the GTV that had similar intensity to it on the CT image. However, the RO was able to distinguish the GTV from neighbouring healthy tissue from prior knowledge of the original planning CT that was enhanced with an intravenous (IV) contrast agent (Sections 5.1 and 5.4.1). This example highlights one of the limitations of voxel intensity-based DIR algorithms where algorithm performance can be affected by external factors not usable by the algorithm. Since all the original pCTs in this study were contrast-enhanced and the subsequent CTs were not, it was difficult to determine how DIR algorithm performance was affected. Although, patient P2's third CT scan (matched to the second CT) did allow a comparison of the DSI scores from contrast-plain CT matches to those from plain-plain CT matches. The plain-plain CT match did have significantly higher DSI scores compared to the contrast-plain CT match (Student's t-test, $p = 0.007$), though this result is for only one patient.

It was also observed that significant proportions of the discrepancies between the DIR-generated contours and the RO-drawn contours occurred at the superior and inferior extremities of the registered VOIs (Figure 5.7 for example). This may be explained by the lower resolution in the axial direction (due to the 3 mm CT slice thickness) in comparison to the image resolution in the coronal and sagittal planes that is higher (≤1 mm).

Qualitative assessment of the deformed images also helped to indicate the likely causes of the anatomical changes in the various VOIs. A visualisation of the DVF in the form of a colour intensity map indicating the magnitude of the deformation vectors in a given dimension was used to show discernible patterns in the DVF that would indicate patient weight loss or swelling of the neck (Section 5.4.2, Figures 5.8 and 5.9). For both examples, the patient records were checked to confirm the validity of these DVF-based speculations. For patient P4 the dietician's notes indicated that the patient lost approximately 7.4 kg in body weight during the 5 weeks of treatment between the original planning CT and the second CT images, which was consistent with the DVF visualised in Figure 5.8. Likewise, the DIR results for patient P3 were consistent with the notes of the Medical Oncologist that swelling of the neck was observed 2 days before the patient was scanned for the second time. Eventually, analysis of DIR results of a larger group of patients to characterise observed phenomena may facilitate early recognition of these effects and can serve as an indicator for adaptation of an on-going treatment plan.

6.5. Summary – Comparative Analysis of Clinical and Phantom Investigations

DIR performance in clinical cases was generally better than in the phantom investigations. Although the phantom studies were controlled experiments, the presence of air gaps in the VOIs (RANDO and CIRS phantoms) prevented the simulation of realistic clinical deformations. In contrast, the clinical studies were realistic but uncontrolled, so the actual deformations could only be evaluated by assuming the RO's delineation of the VOIs was a reliable golden standard. This is not necessarily the case when considering inter-observer and intra-observer variability.

For instance, the Modified Demons algorithm very accurately determined air-tissue interfaces in the clinical cases, especially since all the GTVs in the study were adjacent to the patients' airways (Figure 6.1). This may be related to the fact that the Modified Demons algorithm is based on iso-intensity contours in the image and hence a sharp contrast in voxel intensities becomes an easily detectable barrier of the VOI. In the phantoms, the air gaps meant the voxel intensities within the contour were sharply heterogeneous, thus confounding the algorithm's solution convergence.

In future, the study can be improved in a number of ways: (I) ensuring phantoms have no air gaps in the VOIs. (II) Use *in silico* deformations of phantom and CT images that are controlled and known precisely and accurately. This method has been used in a number of studies and can be a suitable component in a quality assurance protocol of DIR [23, 24]. (III) Use a DIR algorithm that allows control of smoothing parameters and in-built boundary conditions. SA is a proprietary programme and it was not possible to characterise or adjust the boundary conditions that have been implemented.

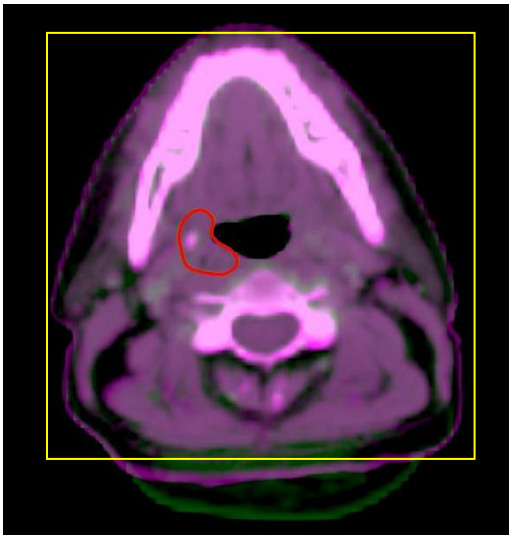

Figure 6.1: Axial slice through deformably registered images of patient P3 shown in colour blend mode. The red contour corresponds to the DIR-propagated GTV. The yellow box represents the ROI used for registration.

In terms of contouring, the phantoms are constructed of distinct compartmentalised structures that are very easily contoured in a short time. The TVs and OARs in clinical cases are intertwined and generally have low intensity gradients that do not lend themselves to easy contouring. Therefore, the overall high accuracy of DIR and the propagation of contours in subsequent images have a significant impact on the time spent by the RO re-contouring the VOIs on a new CT image. In this study, the 3 ROs reported that typically full contouring of all pertinent VOIs in a HNC patient takes approximately 1.5 hours. In comparison, the inspection and occasional minor modification of DIR-generated contours took the ROs 0.5 hour, giving a reduction of ~67% in contouring time. The time to carry out a complete DIR process took 10-15 minutes per patient, which means that DIR contour propagation gives a net reduction of ~47% in contouring time. Chao *et al* reported that DIR was able to save physicians 26-47% in contouring time compared to contouring from *de novo* in a study investigating the impact of automatic contouring on the consistence and time spent for head and neck contouring [56]. Such substantial time savings can have substantial benefits for clinical workflow that alone would justify the use of DIR in HNC radiotherapy.

6.6. Conclusions

In this study a protocol has been designed and validated for the use of a commercial automatic deformable image registration programme in head and neck cancer radiotherapy. It was demonstrated that the Modified Demons algorithm implemented in the SmartAdapt DIR system resulted in clinically acceptable CT image deformations, assessed and approved by expert radiation oncologists and medical physicists. The phantom investigations highlighted the potential impacts of ROI selection and CT image slice thickness on DIR performance, even though the results were of limited applicability to clinical scenarios, due largely to the in-built boundary conditions of the DIR algorithm that suit clinical scenarios. It was also shown that DIR significantly reduced re-contouring time and can thus greatly improve clinical workflow. DIR contour propagation was shown to be of sufficient accuracy for clinical application, provided the results are checked and corrected as necessary by experienced physicians, hence demonstrating the feasibility of using DIR in HNC adaptive radiotherapy pending further investigations.

6.7. Future Work

While the time-saving advantages of DIR alone justify its integration in routine clinical practice of HNC radiotherapy, the application of DIR as part of an AR strategy for head and neck cancer can yield markedly improved treatment outcomes [50, 57-59]. However, before this can be realised as part of routine clinical practice, there are many areas that need further investigation.

Firstly, studies like those by Kirby *et al* [23], Hardcastle *et al* [55], and Fallone *et a* [24] need to be carried out to investigate the suitability of different DIR algorithms to different clinical sites and their different ranges of anatomical variation. Secondly, validation studies of using

inter-modality DIR, like matching planning CTs to routine cone-beam CT (CBCT) images or functional positron emission tomography (PET) and magnetic resonance imaging (MRI) images in achieving both online and offline ART [60, 61].

Thirdly, whether by using deformable phantoms or direct application to retrospective and prospective randomised clinical trials, proof of concept studies are needed to design a method to verify true voxel intensity movement, i.e. a way to verify the algorithm's interpretation of a given change. This is an important limitation when considering dose deformation or dose accumulation in the context of ART.

These related concepts involve the deformation of the TPS-calculated dose delivered to the patient based on the deformed image and the DVF calculated by DIR. While the topic of dose deformation/accumulation has been beyond the scope of this study, it needs to be discussed briefly here, as it is the next step after DIR and has been the subject of considerable debate in the DIR and ART research community [12, 14]. The errors in the DVF found in the RANDO phantom experiment (Section 6.4.1) support the argument against deforming dose because they show that delivered radiation dose cannot be tracked based on the changes in voxel intensities. That is, a voxel with intensity x on the deformed image that is matched to a voxel with the same intensity x on the target image do not necessarily represent the same tissue or portion of anatomy. Therefore, it is argued that dose deformation should not be used clinically until it is done with algorithms that have been validated against measurement, rather than being merely based on image manipulation. Proponents on the other hand, advocate that it is better to attempt an approximation of the change in dose distribution than adapting a treatment plan based only on image registration, and have put forth deformable dosimetric phantom studies that can yield accurate predictions of the dose distribution as examples. However, the results in this study point to agreement with those opposing dose deformations: without verifying DIR accuracy down to voxel-by-voxel agreement, and not

just correlation and similarity measures, it is inappropriate to deform dose in a clinical setting.

Addressing the issues outlined above and other aspects of image guided radiotherapy, like calculating dose using per-fraction cone-beam CT images [62], would then allow for the design of an adaptive radiotherapy strategy for head and neck cancer.

Bibliography

1. Harrison, L.B., R.B. Sessions, and W.K. Hong, eds. *Head and Neck Cancer: A Multidisciplinary Approach*. Third ed. 2009, Lippincott Williams & Wilkins: Philadelphia, PA, USA.
2. Bushberg, J.T., et al., *The Essential Physics of Medical Imaging*. Second ed. 2002, Philadelphia, USA: Lippincott Williams & Wilkins.
3. Seltzer, S.M.*e.a.*, *ICRU Report 83: Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT)*, in *Journal of the ICRU*, S.M. Seltzer, Editor. 2010.
4. Khan, F., M., *The Physics of Radiation Therapy*. Fourth ed. 2010: Lippincott, Williams and Wilkins.
5. Castadot, P., et al., *Adaptive Radiotherapy of Head and Neck Cancer.* Seminars in Radiation Oncology, 2010. **20**: p. 84-93.
6. Yan, D., et al., *Adaptive Radiation Therapy.* Physics in Medicine and Biology, 1997. **42**: p. 123-132.
7. Modersitzki, J., *Numerical Methods for Image Registration*. First ed. 2004, New York: Oxford University Press.
8. Wen, N., et al., *Evaluation of the deformation and corresponding dosimetric implications in prostate cancer treatment.* Physics in Medicine and Biology, 2012. **57**: p. 5361-5379.
9. Tome, W.A. and J.F. Fowler, *On cold spots in tumour subvolumes.* Medical Physics, 2002. **29**(1590-1598).
10. Yan, D., D.A. Jaffray, and J.W. Wong, *A model to accumulate fractionated dose in a deforming organ.* International Journal of Radiation Oncology Biology Physics, 1999. **44**: p. 665-675.
11. Jaffray, D.A., et al., *Accurate accumulation of dose for improved understanding of radiation effects in normal tissues.* International Journal of Radiation Oncology Biology Physics, 2010. **76**: p. S135-S139.
12. Schultheiss, T.E., W.A. Tome, and C.G. Orton, *It is not appropriate to "deform" dose along with deformable image registration in adaptive radiotherapy.* Medical Physics, 2012. **39**(11): p. 6531-6533.
13. Schwartz, D.L. and L. Dong, *Adaptive radiation therapy for head and neck cancer - Can an old goal evolve into a new standard?* Journal of Oncology, 2011. **2011**: p. 1-13.
14. Yeo, U.J., et al., *Is it sensible to "deform" dose? 3D experimental validation of dose-warping.* Medical Physics, 2012. **39**(8): p. 5065-5072.
15. Zhang, T., et al., *Automatic delineation of on-line head and neck computed tomography images: toward on-line adaptive radiotherapy.* International Journal of Radiation Oncology Biology Physics, 2007. **68**(2): p. 522-530.
16. Sims, R., et al., *A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck.* Radiation Therapy and Oncology, 2009. **93**(3): p. 474-8.
17. Young, A., et al., *Atlas-based segmentation improves consistency and decreases time required for contouring postoperative endometrial cancer nodal volumes.* International Journal of Radiation Oncology Biology Physics, 2011. **79**(3): p. 943-947.

18. Castadot, P., et al., *Comparsion of 12 Deformable Registration Strategies in Adaptive Radiation Therapy for the Treatment of Head and Neck Tumors.* Radiation Therapy and Oncology, 2008. **89**: p. 1-12.
19. Samant, S., et al., *High Performance Computing for Deformable Image Registration: Towards a New Paradigm in Adaptive Radiotherapy.* Medical Physics, 2008. **35**(8): p. 3546-3553.
20. Bender, E.T. and W.A. Tome, *The utilization of consistency metrics for error analysis in deformable image registration.* Physics in Medicine and Biology, 2009. **54**: p. 5561-5577.
21. Boswell, S., et al., *Automatic registration of megavoltage to kilovoltage CT images in helical tomotherapy: An evaluation of the setup verification process for the special case of a rigid head phantom.* Medical Physics, 2006. **33**: p. 4395-4404.
22. Yeo, U.J., et al., *Validation of common deformable image registration (DIR) algorithms with a novel deformable phantom*, in *EPSM-ABEC Conference*. 2011, Medical Radiation Physics Group, RMIT University: Darwin, Australia.
23. Kirby, N., et al., *The need for application-based adaptation of deformable image registration.* Medical Physics, 2012. **40**(1): p. 1-9.
24. Fallone, B.G., et al., *Assessment of a commercially available automatic deformable registration system.* Journal of Applied Clinical Medical Physics, 2010. **11**(3).
25. Yang, D., et al., *Technical Note: DIRART - A Software Suite for Deformable Image Registration and Adaptive Radiotherapy Research.* Medical Physics, 2010. **38**(1): p. 67-77.
26. Xing, L., J.V. Siebers, and P.J. Keall, *Computational Challenges for Image-guided Radiation Therapy: Framework and Current Research.* Seminars in Radiation Oncology, 2007. **17**: p. 245-257.
27. McInerney, T. and D. Terzopoulos, *Deformable Models in Medical Image Analysis: a Survey.* Medical Image Analysis, 1996. **1**(2): p. 91-108.
28. Shen, J.-K., B.J. Matuszewski, and L.-K. Shark, *Deformable Image Registration*, in *IEEE Xplore Digital Library*. 2005, Institute of Electronics and Electronics Engineers.
29. Maintz, J.B. and M.A. Viergever, *A survey of medical image registraton.* Medical Image Analysis, 1998. **2**(1): p. 1-36.
30. Mencarelli, A., et al., *Validation of deformable registration in head and neck cancer using analysis of variance.* Medical Physics, 2012. **39**(11): p. 6879-6884.
31. Varian, *Registration and SmartAdapt Reference Guide*. 2010, Varian Medical Systems, Inc.
32. Thirion, J.-P., *Image Matching as a Diffusion Process: An analogy with Maxwell's Demons.* Medical Image Analysis, 1998. **2**(3): p. 243-260.
33. Horn, B.K.P. and B.G. Schunck, *Determining optical flow.* Artificial Intelligence, 1981. **17**: p. 185-203.
34. Yang, D., *DIRART (Deformable Image Registration and Adaptive Radiotherapy) Software Suite*, in *User Instruction Manual*, I. El Naqa, Editor. 2009, Deshan Yang: Washington University in Saint Louis.
35. Lu, W., et al., *Fast free-form deformable registration via calculus of variations.* Physics in Medicine and Biology, 2004. **49**: p. 3067-3087.
36. Iu, S.-L. and Y.-T. Lin, *Re-examining the optical flow constraint - a new optical flow algorithm with outlier rejection*, in *IEEE Xplore Digital Library*. 1999, Institute of Electronics and Electronics Engineers.
37. Wang, H., et al., *Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy.* Physics in Medicine and Biology, 2005. **50**: p. 2887-2905.

38. Kostelec, P.J., J.B. Weaver, and D. Healy, *Multiresolution elastic image registration.* Medical Physics, 1998. **25**(9): p. 1593-1604.
39. Pluim, J.P.W., J.B. Maintz, and M.A. Viergever, *Mutual Information Based Registration of Medical Images: A Survey.* IEEE Transactions on Medical Imaging, 2003. **22**(8): p. 986-1004.
40. Esteban, M.D. and D. Morales, *A Summary of Entropy Statistics.* Kybernetika, 1995. **31**(4): p. 337-346.
41. Roshni, V.S. and K. Revathy, *Using Mutual Information and Cross Correlation as Metrics for Registration of Images.* Journal of Theoretical and Applied Information Technology, 2008. **4**(6): p. 474-481.
42. Dice, L.R., *Measures of the amount of ecologic association between species.* Ecology, 1945. **26**: p. 297-302.
43. Podgorsak, E.B., *RADIATION ONCOLOGY PHYSICS: A HANDBOOK FOR TEACHERS AND STUDENTS*. 2005, Vienna: International Atomic Energy Agency.
44. Yang, D., *DIRART Suite*. 2009, Washington University: St Louis.
45. DRS. *General Principles Associated with Good Imaging Technique: Technical, Clinical and Physical Parameters*. [cited 2012 31/12/2012]; Guidlines on CT imaging quality from the Danish Society of Radiology].
46. *RANDO® Phantom Technical Data Sheet*, T.P. Library, Editor. 2012, The phantom Laboratory: New York.
47. *IMRT Head and Neck Phantom*, C.I.R.S. Inc., Editor. 2011, CIRS: Norfolk, Virginia.
48. National_Institute_of_Standards_and_Technology, *The NIST Reference on Constants, Units and Uncertainty*, in *Uncertainty of Measurement Results*. 2011, Physical Measurement Laboratory: Online Website: http://physics.nist.gov/cuu/Uncertainty/combination.html.
49. *Clinical protocol - head and neck treatment using RapidArc technique*, in *Blood & Cancer Centre, Wellington Regional Hospital*. 2012. p. 1-10.
50. Schulze, D., et al., *Comparison of Various Online IGRT Strategies: The Benefits of Online Treatment Plan Re-optimization.* Radiotherapy Oncology, 2009. **90**: p. 367-376.
51. Berthelet, E., et al., *CT slice index and thickness: Impact on organ contouring in radiation treatment planning for prostate cancer.* Journal of Applied Clinical Medical Physics, 2003. **4**(4): p. 365-373.
52. Tunio, M.A., et al., *Influence of CT slice thickness on volume and dose uncertainty for different organs during treatment planning for early prostate cancer.* Pakistan Journal of Radiology, 2012. **20**(1): p. 87-91.
53. Grégoire, V., et al., *CT-based delineation of lymph node levels and related CTVs in the node-negative neck: DAHANCA, EORTC, GORTEC, NCIC, RTOG consensus guidelines.* Radiotherapy Oncology, 2003. **69**(3): p. 227-36.
54. Brouwer, C.L., et al., *3D Variation in delineation of head and neck organs at risk.* Radiation Oncology, 2012. **7**(32): p. 1-9.
55. Hardcastle, N., et al., *A multi-institution evaluation of deformable image registration algorithms for automatic organ delineation in adaptive head and neck radiotherapy.* Radiation Oncology, 2012. **7**(90).
56. Chao, K.S., et al., *Reduce in variation and improve efficiency of target volume delineation by a computer-assisted system using a deformable image registration approach.* International Journal of Radiation Oncology Biology Physics, 2007. **68**(5): p. 1512-21.
57. Schwartz, D.L., et al., *Adaptive radiotherapy for head-and-neck cancer: initial clinical outcomes from a prospective trial.* International Journal of Radiation Oncology Biology Physics, 2012. **83**(3): p. 986-93.

58. Grégoire, V., et al., *Radiotherapy for head and neck tumours in 2012 and beyond: conformal, tailored, and adaptive?* The Lancet Oncology, 2012. **13**(7): p. 292-300.
59. Schwartz, D.L., et al., *Adaptive radiotherapy for head and neck cancer-Dosimetric results from a prospective clinical trial.* Radiotherapy Oncology, 2013. **106**(1): p. 80-84.
60. Elstrom, U.V., et al., *Daily kV cone-beam CT and deformable image registration as a method for studying dosimetric consequences of anatomic changes in adaptive IMRT of head and neck cancer.* Acta Oncologica, 2010. **49**: p. 1101-1108.
61. Söhn, M., et al., *Model-independent, multimodality deformable image registration by local matching of anatomical features and minimization of elastic energy.* Medical Physics, 2008. **35**(3): p. 866-78.
62. Hatton, J., B. McCurdy, and P.B. Greer, *Cone Beam Computerized Tomography: The Effect of Calibraiton of the Hounsfield Unit Number to Electrong Density on Dose Calculation Accuracy for Adaptive Radiation Therapy.* Physics in Medicine and Biology, 2009. **54**: p. N329-N346.